Article, Author's Postprint

Mohammadi, Farid; Kopmann, Rebekka; Guthke, Anneli; Oladyshkin, Sergey; Nowak, Wolfgang

Bayesian selection of hydro-morphodynamic models under computational time constraints

Advances in Water Resources

Bayesian selection of hydro-morphodynamic models under computational time constraints

Farid Mohammadi [a]

Rebekka Kopmann [b]

Anneli Guthke [a]

Sergey Oladyshkin [a,*]

Wolfgang Nowak [a]

[a] Department of Stochastic Simulation and Safety Research for Hydrosystems (IWS/SC SimTech), University of Stuttgart, Germany

[b] Federal Waterways Engineering and Research Institute (BAW), Karlsruhe, Germany

[*] Corresponding author

Abstract:

A variety of empirical formulas to predict river bed evolution with hydro-morphodynamic river models exists. Modelers lack objective guidance of how to select the most appropriate one for a specific application. Such guidance can be provided by Bayesian model selection (BMS). Its applicability is however limited by high computational costs. To transfer it to computationally expensive river modeling tasks, we propose to combine BMS with model reduction based on arbitrary Polynomial Chaos Expansion. To account for approximation errors in the reduced models, we introduce a novel correction factor that yields a reliable model ranking even under strong computational time constraints. We demonstrate our proposed approach on a case study for a 10-km stretch of the lower Rhine river. The correction factor may shield us from misleading model ranking results. In our case, the correction factor was shown to increase the confidence in model selection.

Keywords:

Hydro-morphodynamics

Bayesian model selection

Bayesian model evidence

Reduced model

Polynomial chaos expansion

Surrogate-enabled model selection

1   Introduction

Modern river management must consider numerous requirements, such as flood protection, provision of safe and efficient navigation, floodplain agriculture, ecology and recreation. These requirements are often in conflict with each other, especially for multi-function rivers in densely populated areas. As the hydraulic regime, sediment transport and riverbed elevation play a significant role in handling these requirements, a good grasp of hydro-morphodynamic processes is important.

The river hydrodynamics can be described by the conservation laws of mass, energy, and momentum applied to water as an incompressible fluid. Hydrodynamic models enable a river engineer to investigate the flow processes in water bodies and provide predictions at various scales. Water flowing in natural rivers is capable of scouring riverbeds, carrying particles (heavier than water) and depositing materials, resulting in sediment transport and corresponding changes in bed topography. The understanding of river morphodynamics is essential for proper management and prediction of critical erosion and sedimentation processes occurring in streams, such as bank erosion, sediment wave propagation, interactions with man-made structures (bridges, weirs), silting up of reservoirs, sediment mining, degradation/aggradation of river channels and planform changes.

However, sediment transport formulas (bed load or total load) have improved only insignificantly over time, and the proposed theories have not been able to predict the actual sediment discharges accurately. Sediment transport depends, in a complex and poorly understood way, on several parameters related to flow and sediment (Beckers et al., 2017). Therefore, almost all the existing sediment transport predictors have been derived under certain simplifications and assumptions. These are responsible for a large fraction of inaccuracies in sediment estimations and seldomly aligned with natural processes in rivers (Habibi, 1994; Papanicolaou et al., 2008). Another difficulty in developing and assessing sediment transport formulas is the lack of sediment data. The amount of available sediment data cannot fully represent the temporal and spatial variations in natural rivers.

Most sediment transport formulas have been developed in two stages. First, a theoretic relationship is postulated which attempts to relate the sediment transport rate to hydraulic and sedimentological quantities. Second, the proposed formula is revised or parameters are adjusted after comparison with data (e.g. Meyer-Peter et al., 1934 or Ackers and White, 1973). Numerous sediment transport formulas originate from experimental data by the original authors (e.g. Schoklitsch, 1934), which usually have been accompanied by data obtained by one or several previous investigators, commonly including that of Gilbert and Murphy (1914). Moreover, the capability of a formula has been regularly evaluated based on its agreement with a limited amount of field data (Müller, 1937). Altogether, most formulas owe their origin to a relatively limited data set, and their effectiveness has been tested via comparatively few field data. Consequently, no single formula has been either widely accepted or recognized as being particularly appropriate.

While the scarcity of reliable field data has apparently not been helpful to validate sediment transport formulas, there have been some attempts to evaluate and compare their performances. Since the first comparison study by Johnson (1943), a long list of further evaluation studies have been presented in the literature (e.g., Cole et al., 1973; Hubbell and Matejka, 1959; Jordan, 1965; Martin and Ham, 2005; Yang and Wan, 1991). Batalla (1997) concludes that the degree of accuracy between observed and predicted values differs significantly from one formula to another. Similarly, most evaluations deduce that no formula performs reliably across many different situations; this can be accredited to limitations of test data and to restrictions of experiments and the physical characteristics of the transport phenomena (López et al., 2014). Testing the predictive abilities of these formulas over intermediate temporal scales (approximately decades) and spatial scales (a length of at least several kilometers) is limited by a lack of field data on sediment transport rates at these levels, especially due to the spatiotemporal variability of sediment transport and riverbed composition (Wilcock, 2001). According to Gomez and Church (1989), there are more sediment transport formulas than reliable data to test them. However, the studies cited above demonstrate that no individual formula can be labeled as ideal for the full range of existing morphodynamic conditions.

Regardless of the enhanced research in the field of sediment transport shortly discussed above, the greatest challenge in many hydro-morphodynamic studies is to determine the best formula that describes the morphodynamic condition of a given river stretch. Therefore, a set of competing formulas should be considered and objectively ranked in the light of the specific characteristics and available data of the study site. Evaluating the relative merits of competing formulas can be challenging in the face of limited data and uncertain model parameters and inputs. Hence, there is a need for a consistent framework that allows to rank competing hydro-morphodynamic models as a basis for objective model choice. This framework must account for the uncertainty in model choice, in parameters and in data that plagues most hydro-morphodynamic situations.

For this task, we suggest to apply the framework of Bayesian model selection to identify the most suitable hydro-morphodynamic model. Bayesian model selection (BMS) uses rigorous concepts of probability theory and numerical integration to obtain a model ranking (Draper, 1995; Raftery, 1995). Objective choice of models based on this ranking significantly reduces the potential bias in modeling. However, direct implementation of the BMS framework requires a large number of model runs and becomes computationally very demanding (Schöniger et al., 2014). Several approaches to reduce the computational burden of BMS have therefore been proposed in the field of water resources, e.g. thermodynamic integration (Liu et al., 2016), nested sampling (Skilling et al., 2006) or Gaussian mixture importance sampling (Volpi et al., 2017), to name a few selected examples. These approaches rely on efficient numerical solutions to determine Bayesian model weights, but still require on the order of tens to hundreds of thousands of model runs.

Generally, Bayesian inference theory can be seen as a statistically rigorous approach to deal with uncertainty. It provides probabilistic information on the remaining uncertainty in parameters and predictions after calibration. Camacho et al. (2015) provide a comparison of traditional Bayesian approaches such as Bayesian Monte Carlo and Markov chain Monte Carlo with the Generalized Likelihood Uncertainty Estimation for uncertainty analysis in hydraulic and hydrodynamic modeling. Bayesian updating of parameters for a sediment entrainment model via Markov chain Monte Carlo

has been presented in Wu and Chen (2009). However, sampling-based stochastic approaches are more expensive than classical optimization-based (deterministic) calibration techniques, because they need to explore the full range of possible model outcomes for the entire feasible parameter range requiring many model runs. A particular reason for the high computational costs is that, when exploring the full range of possible model outcomes, many model runs are rejected upon comparison with data. Therefore, the high computational costs of forward modeling get multiplied by large factors in stochastic calibration methods. Recently, various model reduction techniques have been combined with Bayesian updating to reduce the computational burden (see, e.g., Bazargan et al., 2015; Jin, 2008; Laloy et al., 2013; Zeng et al., 2016). For a comprehensive review of approaches to model reduction in water resources, we refer to Razavi et al. (2012). A model calibration framework using strict Bayesian principles combined with the so-called arbitrary polynomial chaos expansion for model reduction has been established in Oladyshkin et al. (2013a). This approach was the first to be applied successfully to a real large-scale pilot site for geological carbon dioxide storage in Europe (see Oladyshkin et al., 2013b). The specific limitations of polynomial chaos-based methods for Bayesian inversion have been discussed in Lu et al. (2015). Several attempts to reduce full-complexity models in river engineering have been discussed in Nicholas (2009) and Liang et al. (2015), however not in a Bayesian context.

We continue this recent trend toward stochastic calibration based on so-called response surfaces, reduced models, surrogate models, emulators, meta-models, etc. To accelerate the forward modeling, we will construct response surfaces that represent the dependence of the outputs of full-complexity hydro-morphodynamic models on the modeling parameters. However, due to computational time constraints, the response surface might not be trained well enough to represent the full-complexity model sufficiently well. This indicates that there is a trade-off between a wide context of applicability and large computational time. We suggest to use iterative improvement (see Li and Marzouk, 2014; Oladyshkin et al., 2013a) of the response surface within the limited simulation time budget to increase the reliability of the reduced models.

Then, we make use of the fact that reduced models are already established for stochastic calibration and propose an approach to combine BMS with model reduction (Section 3). This combination allows us to transfer BMS to the practical field of river modeling, despite the large computational burden of hydro-morphodynamic models. Previous approaches to reduce the number of required model runs such as efficient numerical integration as described above are not applicable because they cannot reasonably operate on sample sizes on the order of 10–100, which is the computational limit set to our case study. Here, one full-complexity model run takes about 7.5 cluster hours. After construction of the reduced models with this very small number of full-complexity model runs, computational time of the reduced models is negligible (milliseconds).

Since reduced models are only a (possibly rough) approximation of the full-complexity model, the validity of model ranking results based on reduced models clearly depends on the quality of approximation. We argue that the model ranking may turn out differently for reduced models than for full-complexity models, which could trigger biased conclusions and finally possibly adverse management decisions. To avoid such misleading results, we propose to explicitly account for the quality of approximation by the reduced model in model ranking. By applying a correction factor that penalizes

approximation errors, we obtain model ranking results that are more representative of the actual full-complexity model ranking. While BMS has recently gained popularity in various fields of geosciences and engineering to deal with model choice uncertainty (e.g., Brunetti et al., 2017; Elsheikh et al., 2012; Lötgering-Lin et al., 2016; Marshall et al., 2005; Poeter and Anderson, 2005; Refsgaard et al., 2012; Schöniger et al., 2015; Wöhling et al., 2015), applications to reduced models to circumvent computational limitations are a logical next step but only start to emerge. For example, Jia et al. (2018) estimates integrated model uncertainty for 12 alternative carbon dioxide storage models using Bayesian model averaging approach and polynomial chaos expansion. Due to the computational challenging physical problem, Jia et al. (2018) suggests to estimate the corresponding model weights using the reduced model obtained from 50 randomly sampled model runs. Targeting the same direction, Xue et al. (2018) have employed probabilistic collocation method based on prior distribution to construct polynomial chaos expansion for a groundwater flow application. Then they have used the obtained reduce model to produce a model averaged prediction based on Bayesian model weights. Their comparison with traditional Monte Carlo implementation for the full models indicates that reduced models can yield a very accurate model weighting. However, they relied on more than 1000 model runs to establish six reduced models. In typical practical applications, such a comparison with a reference solution for model weights is infeasible, given that a reliable estimation of model weights requires on the order of millions of model runs (Schöniger et al., 2015; 2014). To the best of our knowledge, we are the first to propose a correction factor that establishes a link between the model ranking results for reduced models and for full-complexity models, working with just the number of model runs required for building the reduced models. With the proposed approach, we hope to promote the use of BMS in river engineering and in fact any discipline that is confronted with model choice among computationally demanding models.

In short, the main contributions of our study are: (1) applying an advanced model reduction strategy to hydro-morphodynamic models, (2) demonstrating Bayesian selection of several alternative hydro-morphodynamic models based on model reduction, and (3) correcting the obtained model ranking results in the light of the approximation errors made by the reduced models.

We demonstrate the applicability of our proposed Bayesian concept to river engineering in both a synthetic benchmark study and a real-data case study of the Lower Rhine in North Rhine-Westphalia, Germany. The models used in our studies are described in Section 2. In Section 3.1, we explain the basic concepts of Bayesian updating and BMS. To satisfy the computational budget constraints of the presented project (limiting ourselves to less than 100 model runs in total), we first build surrogate models (response surfaces) via arbitrary Polynomial Chaos (aPC) (Section 3.2), and then apply iterative Bayesian updating and finally BMS to the reduced models (Section 3.3). The proposed correction factor for surrogate-enabled BMS is introduced in Section 3.4. Details with respect to the implementation of the case studies are given in Section 3.5. We discuss the results of model calibration and selection for both the synthetic benchmark test and the real-data test case in Section 4 and finally draw conclusions in Section 5.

2 Hydro-morphodynamic model

The applied hydro-morphodynamic models have been built in the TELEMAC-MASCARET simulation software, which is an open-source integrated suite of solvers for simulation of free-surface river flow (Hervouet, 2000) (www.opentelemac.org). We use the hydrodynamic module TELEMAC-2D and the morphodynamic module SISYPHE to predict sediment transport processes and the resulting changes in the riverbed. These two modules are internally coupled, i.e. the relevant hydrodynamic variables are exchanged between the hydrodynamic and the morphodynamic modules. We describe the governing equations of the hydrodynamic model and the main considered processes of the morphodnymic model including the applied sediment transport formulas in Section 2.1. Details on the model setup for our specific case study are given in Section 2.2. Further details about the TELEMAC-2D and SISYPHE modules can be found in Hervouet (2007), Tassi and Villaret (2014), and Villaret et al. (2013), respectively.

2.1 Governing equations

To describe the hydrodynamics of the river, TELEMAC-2D solves the shallow water equations, also known as Saint Venant equations:

$$\begin{cases} \dfrac{\partial h}{\partial t} + u\nabla_x h + h\nabla v = S_h \, , \\ \dfrac{\partial u}{\partial t} + u\nabla u = -g\dfrac{\partial Z}{\partial x} + S_x + \dfrac{1}{h}\nabla(h\vartheta_t\nabla u) \, , \\ \dfrac{\partial v}{\partial t} + v\nabla v = -g\dfrac{\partial Z}{\partial y} + S_y + \dfrac{1}{h}\nabla(h\vartheta_t\nabla v) \, , \end{cases} \qquad (1)$$

where h [m] denotes water depth, u and v [m/s] are flow velocities in x- and y-direction, respectively, Z [m] is the free surface elevation, and S_h [m/s] is a source or sink term of the fluid. S_x and S_y [m/s^2] represent deceleration due to bed roughness, and ϑ_t [m/s^2] is the viscosity. For sediment transport processes, the bed shear stress plays a major role. Local rates of sediment transport are influenced by near-bed flow fields. These flow fields are mostly governed by the river bed topography and bed evolution. To predict bed evolution, the morphodynamic module (SISYPHE) solves the Exner equation:

$$(1-n)\frac{\partial Z_f}{\partial x} + \nabla Q_b = 0 \, , \qquad (2)$$

where n [-] is the non-cohesive bed porosity, Z_f [m] is the bed level, and Q_b [m^2 s^{-1}] is the volume of sediment transport per unit width. The new river bed is used in calculation of the velocities and water depth in Eq. (1). The mentioned procedure is repeated according to the prescribed coupling time between hydrodynamic and morphodynamic modules.

SISYPHE provides an extensive library of sediment transport relations to choose from and the opportunity of own implementation of any other formula. In this study, we have applied the six

following formulas that are widely used: the Meyer-Peter and Müller formula (Meyer-Peter and Müller, 1948), the Einstein-Brown formula (Einstein et al., 1950), the Modified Engelund-Hansen formula by Chollet and Cunge (Chollet and Cunge, 1980; Engelund and Hansen, 1967), the van Rijn formula (van Rijn, 1984), the Hunziker formula (Hunziker, 1995), and the Wu formula (Wu et al., 2000).

The Meyer-Peter and Müller formula is probably the most commonly used in this field, but developed assuming a uniform sediment distribution as all of the older formulas. The Hunziker and Wu formulas were both explicitly developed for graded sediment and promise a better behavior in case of non-uniform sediments. Here, a multi-grain approach was used considering different sediment classes. For each sediment class, the sediment discharge is calculated separately by the sediment transport formula.

All of the sediment transport formulas include one or more highly intercorrelated hydrodynamic variables (depth averaged flow velocity, water depth, energy gradient as water surface slope and water discharge). Therefore, they can be regarded as scaling equations of the flow (Gomez and Church, 1989). Overall, this might be a reason why so many variations have been presented so far, why judging their merits remains challenging, the physics of sediment transport remain incompletely understood, why there is no good reason to a priori favor any of the existing formulas over others, and none of existing formulas offers a complete representation of the natural hydro-morphodynamic system. Instead, the most suitable sediment transport formula for a specific modeling task needs to be chosen carefully in the light of the system settings and the measured data. In this study, we will assess these six competing formulas using the rigorous Bayesian framework for model selection.

Furthermore, the calculated sediment discharge will be changed by additional formulations to consider the following effects. A slanting riverbed leads to an increase of the sediment transport rate in the downslope direction, and to a reduction of the sediment rate in the upslope direction. This so-called slope effect changes the sediment rate and its direction. In this study, the modification of magnitude was calculated with the (Koch and Flokstra, 1980) formula and the direction modification with the (Talmon et al., 1995) formula. Additionally, the sediment movement direction is also influenced by the helical flow or also known as secondary current effect in river bends. This is accounted for with the expression supposed by Engelund (1974).

2.2 Model setup

An existing hydro-morphodynamic model of a 10 km long stretch in the central reach of the Lower Rhine at Düsseldorf/Neuss (Rhine-km 739.0 to 749.0) was chosen as case study to illustrate BMS of sediment transport formulas. This section belongs to the free flowing part and is characterized by intense bends and active sediment transport. Dumping and dredging activities were conducted for a long time (Messing et al., 2008) and provide safe inland navigation. The river stretch has an average slope of about 0.17‰, and is about 400 m wide. The model geometry was created using a Digital Terrain Model (DTM), which included a comprehensive database on the main river channel, the floodplains, as well as groynes, longitudinal and bridge structures in the study area. A two-dimensional numerical morphodynamic transport model was provided by the Federal Waterways Engineering and Research Institute (BAW) in Karlsruhe, Germany (Backhaus et al., 2014; Riesterer, 2015;

Wenka et al., 2015). The numerical model of the 10 km Rhine section has an unstructured, triangulated grid with 41,277 nodes and 81,684 elements. The average node distance is about 10–15 m in the main channel and 50 m in the floodplains. At relevant hydraulic structures, the node distance decreases to 5–10 m. In the 500 m inlet region, the groynes were removed from the model to smooth the incoming flow and avoid numerical instabilities. At the inlet and outlet boundaries, the whole cross-section was considered for the hydraulics (TELEMAC-2D). For the morphodynamics (SISYPHE) only the part of the main channel was opened as it is assumed that relevant sediment transport only occurs in this section. To obtain initial conditions for the hydrodynamics, a steady-state calculation was carried out first with the described boundary conditions for the initial time. For the morphodynamics and sediment transport, the grain size distribution was discretized by 10 sediment classes. A mean grain size over the whole model area was used as initial condition for the morphodynamics. Different roughness zones with their own roughness values determine the natural flow resistance in the model area. The five zones categorize the domain of the river stretch: the main channel, the floodplain, the groynes, the vegetation area and the residence area. As boundary conditions, time dependent discharges were set as the inlet and corresponding water levels at the outlet. From the gauge Düsseldorf (Rhine-km 744.2) located inside the model area, both values were taken and adapted to the location of the model boundaries. For the morphodynamics, an equilibrium condition was posted at the inlet boundary. Therefore, the sediment inflow was calculated according to a fixed bottom level at the entrance. The considered hydrograph starts on 27th of June 2002 and ends on fifth of May, 2004.

Since the model of the 10 km river stretch required a considerably large computational time, the concept of morphological factor was used to reduce computation time. Therefore, the hydrograph was compressed in time by the morphological factor while the results of the morphological simulation were enlarged by the same factor. This procedure reduces the computation time about the morphological factor (Wurms and Schröder, 2012). But even with a factor of 4, the computation time was approximately 7.5 h on the BAW parallel cluster "brutus" (Bull Blade X system) using 60 cores.

2.3   Parameterization

Five parameters (mean grain size, two of the five roughness coefficients, one parameter for the magnitude of the slope effect and one for the effect of secondary currents) were declared as calibration parameters. Their initial values came from the original model, which was manually calibrated at BAW using the Meyer-Peter and Müller formula (see Table 1). Typically, the mean grain size and the roughness coefficient of the main channel are the most sensitive parameters in morphodynamic simulations of rivers. Changing the mean diameter will change the grain size of each sediment class proportionally. Furthermore, the roughness of the floodplains in case of flood events and the slope and secondary currents effect in this river stretch with intense bends are significant for calibration in this case.

As target quantities for calculation, we chose the spatially averaged bottom level, because the evolution of the riverbed is a key quantity of interest in morphodynamic simulations. The averaged evolution is

Table 1:   The selected uncertain parameters and their ranges.

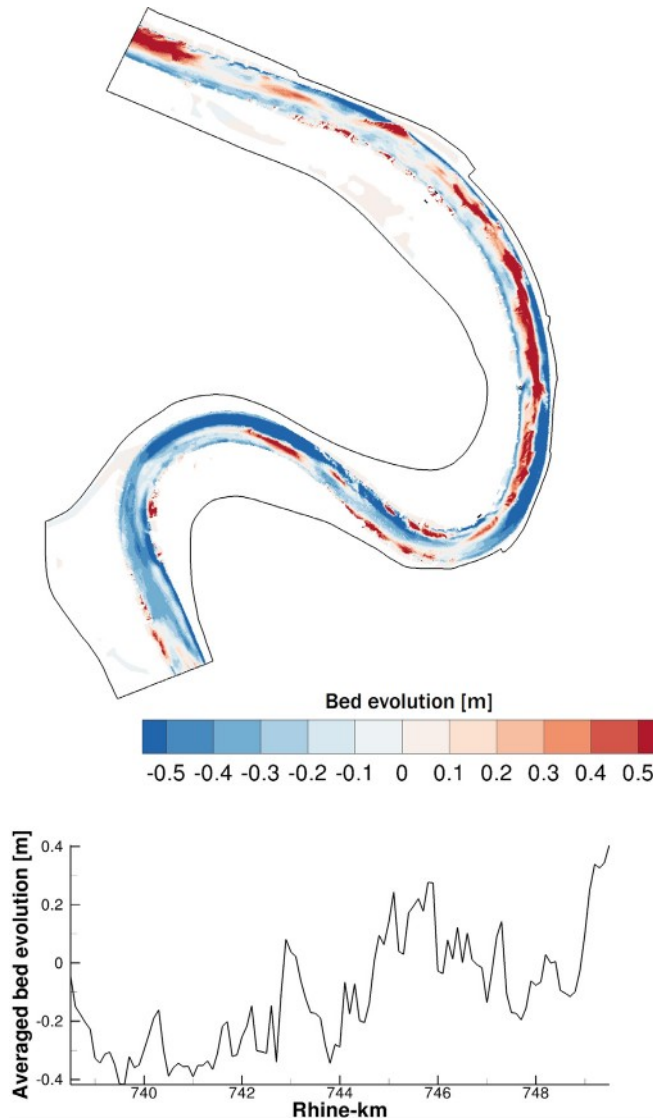| Uncertain parameter's name and unit	Lower bound	Upper bound
Mean sediment diameter [m]	0.01856	0.02051
Roughness coefficient of the main channel [m]	0.49	0.51
Roughness coefficient of the floodplain [m]	0.098	0.102
Slope effect parameter beta [-]	0.85	4.00
Secondary currents alpha coefficient [-]	0.75	1.00

Fig. 1:   The simulated bed evolution from 2002 to 2004 for the lower Rhine model obtained using the Meyer-Peter and Müller formula (top) and the corresponding averaged bottom evolution along the river stretch (bottom).

chosen, because the local bed level is very much influenced by random processes like e.g. dunes or ripples formations. Additionally, the accuracy of bed level measurements are about +/− 10 cm. In our case, each observation point for the calibration procedure is located at the center of an area with a length of 100 m and with a width equal to the sediment active width of the river stretch. Fig. 1 illustrates the model area and the simulated bed evolution from 2002 to 2004 (top), exemplarily shown

for the Meyer-Peter and Müller sediment transport relation and the corresponding averaged bottom evolution along the river stretch (bottom).

3   Bayesian model selection framework

3.1   Bayesian parameter inference and model selection

In the field of river engineering, typically an iterative manual approach of trial and error is implemented to adjust the simulation models to reproduce past measurements reasonably well. However, such a manual process is inefficient, might lead to suboptimal results and suffers from non-uniqueness (Oliver et al., 2008), i.e. various simulation models and parameter sets can fit the observation data similarly well.

Instead of searching for a single best model and parameter set, we perform parameter inference and model selection within a Bayesian probabilistic framework. In the Bayesian framework, initial knowledge on parameters is encoded in a so-called prior probability distribution. After Bayesian parameter inference (which replaces the act of calibration), one obtains a so-called posterior probability distribution of the parameters, which is typically narrower than the prior distributions.

The posterior parameter distribution $p(\boldsymbol{\omega}|M_k, y_o)$ of parameter $\boldsymbol{\omega}$ within a model M_k is obtained by updating the prior parameter distribution $p(\boldsymbol{\omega}|M_k)$ in the light of observed data y_o according to Bayes' theorem:

$$p(\boldsymbol{\omega}|M_k, \mathbf{y}_o) = \frac{p(\mathbf{y}_o|M_k, \boldsymbol{\omega})p(\boldsymbol{\omega}|M_k)}{p(\mathbf{y}_o|M_k)}. \tag{3}$$

The term $p(y_o|M_k, \boldsymbol{\omega})$ is called the likelihood function and quantifies how well the model predictions $Y_k(\boldsymbol{\omega})$ based on specific parameter combinations match the observed data y_o. $p(y_o|M_k)$ is a model-specific normalizing constant for the posterior distribution of the parameters $\boldsymbol{\omega}$. Assuming independent and Gaussian distributed measurement errors, the likelihood function $p(y_o|M_k, \boldsymbol{\omega})$ is defined as:

$$p(\mathbf{y}_o|M_k, \boldsymbol{\omega}) = (2\pi)^{-N_s/2}|\mathbf{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_o - \mathbf{Y}_k(\boldsymbol{\omega})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}_o - \mathbf{Y}_k(\boldsymbol{\omega})\right)\right], \tag{4}$$

where $\mathbf{R}$ is the diagonal (co)variance matrix of size $N_s \times N_s$ (N_s refers to the length of the observation data set) of measurement error ϵ. Though other arbitrary error distributions could be considered, this specific assumption on the measurement error distribution remains throughout the study.

In parallel to parameter updating, we suggest employing Bayes' theorem a second time on the level of model probabilities to rank the plausibility of competing models in the light of data y_o. The prior belief about a model's adequacy is updated to posterior model probabilities in view of the mismatch between model predictions and observed data. The posterior model probabilities can be used as

weights for model ranking and selection. BMS implicitly follows the principle of parsimony or Occam's razor, such that the posterior model weights reflect an optimal trade-off between goodness-of-fit and model complexity (Gull, 1988).

According to Bayes' theorem, the model weights for N_m competing models M_k can be calculated by:

$$P(M_k|\mathbf{y}_o) = \frac{p(\mathbf{y}_o|M_k)P(M_k)}{\sum_{i=1}^{N_m} p(\mathbf{y}_o|M_i)P(M_i)} \, , \tag{5}$$

in which $P(M_k)$ denotes the prior subjective credibility of model M_k before considering any measured data, $p(y_o|M_k)$ is the likelihood of the model M_k to have produced the data set y_o, and $P(M_k|y_o)$ denotes the posterior (updated) model weight. In this study, we assume equally likely priors of $P(M_k) = 1/N_m$, due to the arguments summarized in Section 1. The denominator in (5) normalizes all model weights to sum up to unity. Hence, the decisive factor for model ranking is the likelihood term $p(y_o|M_k)$ which is called Bayesian model evidence (BME). BME is sometimes also referred to as marginal likelihood because it quantifies the likelihood of the model to have produced the observed data averaged over the complete prior parameter space (Kass and Raftery, 1995):

$$p(\mathbf{y}_o|M_k) = \int_\Omega p(\mathbf{y}_o|M_k, \boldsymbol{\omega})p(\boldsymbol{\omega}|M_k)d\boldsymbol{\omega} \, , \tag{6}$$

where $\boldsymbol{\omega}$ denotes the vector of parameters of model M_k with a dimension equal to the number of parameters $N_{(p, k)}$, Ω is the corresponding parameter space, and $p(\boldsymbol{\omega}|M_k)$ represents the prior distribution of parameters in each model M_k.

BME can be evaluated by integration over the full parameter space Ω (Eq. (6)), referred to as Bayesian integral (Kass and Raftery, 1995). Several approaches exist for solving or approximating Eq. (6). For a comprehensive review of methods, we refer to Schöniger et al. (2014). Here, we use simple Monte Carlo integration (Hammersley, 1960). This approach evaluates the integrand at randomly chosen points in parameter space, i.e. they are randomly drawn from their prior distribution $p(\boldsymbol{\omega}|M_k)$. The Bayesian integral is then determined as the mean value of the evaluated likelihoods calculated by Eq. (4), which is why it is often referred to as the arithmetic mean approach.

The significance in differences between BME values of two competing models can be judged with the help of the rule of thumb proposed by Jeffreys (1967). This rule of thumb applies to the so-called Bayes factor, which is the ratio of model evidences of two models M_1 and M_2, $p(y_o|M_1)/p(y_o|M_2)$. According to Jeffreys (1967), a Bayes factor of 1–3 is not worth more than a bare mention, a factor of up to 10 means substantial evidence in favor of model M_1, a factor between 10 and 100 means strong evidence and a Bayes factor larger than 100 is said to be decisive enough to reject model M_2 against M_1. While these thresholds may serve as guidance, the responsibility of defining an application-specific confidence level at which to reject a model or to choose a single best one out of the set remains with the modeler.

3.2 Construction of reduced models

This study focuses on model selection for computationally challenging simulations. In this case, the model selection analysis only becomes feasible in combination with model reduction, because the integral in Eq. (6) is too expensive to be evaluated otherwise. The same holds true for the parameter updating step according to Eq. (3). Here, model reduction is established using so-called response surfaces. We will construct a response surface via the Polynomial Chaos Expansion (PCE) theory (Wiener, 1938), which shows the dependence of a full-complexity model's output on the modeling parameters via projections of the model function $Y_k(\boldsymbol{\omega})$ onto an orthonormal polynomial basis. PCE offers an accurate high-order method of accounting for nonlinear effects in stochastic analysis. It can be seen as an advanced approach to statistical regression to exhibit the dependence of a model on its parameters (Li and Zhang, 2007).

Conventional PCE methods postulate a correct knowledge of the probability density functions of all input parameters and they are optimal only if they are applied to a small number of certain parametric probability distributions. However, the information regarding the statistical distribution of input parameters is scarce, especially in environmental engineering and sciences (Red-Horse and Benjamin, 2004). Moreover, another source of errors is the subjectivity when building probability density functions of any particular shape from samples of limited size or scattered information. Given the mentioned challenges, (Oladyshkin and Nowak, 2012) presented the generalization of PCE to the arbitrary Polynomial Chaos (aPC) expansion that accommodates a broad spectrum of arbitrary distributions with arbitrary probability measures. The distributions can be either discrete, continuous, or discretized continuous, and they can be defined either analytically (as probability density/cumulative distribution functions), numerically (as histograms) or as raw data sets. The aPC adjusts to arbitrary probability distribution shapes of input parameters and can even work with limited data (Oladyshkin et al., 2011; Oladyshkin and Nowak, 2012). However, incomplete statistical information limits the utility of high-order polynomial chaos expansions (Oladyshkin and Nowak, 2018).

We will consider the full-complexity model M_k that depends on multi-dimensional parameter input $\boldsymbol{\omega}$, i.e $\boldsymbol{\omega} = \{\omega_1, \omega_2, ..., \omega_N\}$ and produces model output $\mathbf{Y}_k = f(\boldsymbol{\omega})$. The subscript k indicates a specific model among a set of competing models, with N being equal to the total number of input parameters. We will replace the full-complexity model M_k by a reduced model $\widetilde{M}_k$, i.e. the model output Y_k will be approximated via a multivariate polynomial (response surface) $\tilde{Y}_k$ as follows:

$$\mathbf{Y}(x, y, t, \boldsymbol{\omega}) \approx \tilde{\mathbf{Y}}_k(x, y, t, \boldsymbol{\omega}) = \sum_{i=1}^{P} \mathbf{c}_i(x, y, t)\phi_i(\boldsymbol{\omega}) . \tag{7}$$

Here, $\phi_i(\boldsymbol{\omega})$ are polynomials that form a multi-variate orthogonal polynomial basis, determined according to the procedure presented by Oladyshkin and Nowak (2012). The values P of used polynomials in Eq. (7) depends on the total number of input parameters N and on the order of the expansion d according to the combinatory equation $P = (N + d)!/(N!d!)$. The coefficients $\mathbf{c}_i$ determine the dependence of the model output $\boldsymbol{Y}_k$ on the input parameters $\omega_1, \omega_2, ..., \omega_N$. As the dependence of model output on parameter values may vary within the domain and over time, the coefficients $\mathbf{c}_i(x, y, t)$

depend on space and time, i.e. they have to be determined for each numerical model cell and time step separately.

In order to compute the coefficients c_i(x, y, t) in the expansion (7), we will employ the non-intrusive Probabilistic Collocation Method (PCM), because it is easy to implement and does not require any rearrangement of the governing equations (Li and Zhang, 2007). The main idea of the PCM method is to identify the unknown coefficients c_i (x, y, t) from the model outputs (x, y, t, ω_j) evaluated with different parameters sets ω_j by curve fitting techniques. A given N-dimensional set of parameters ω_j is named as a collocation point in the multidimensional parameter space. The number of collocation points ω_j is equal to the number P of unknown coefficients c_i in the formulated chaos expansion in (7). Formally, the main idea of the PCM method can be written in the following form:

$$\mathbf{\Phi}(\omega)\mathbf{V}_c(x, y, t) = \mathbf{V}_Y(x, y, t) , \tag{8}$$

where V_c is the $P \times 1$ vector of unknown coefficients c_i, V_Y is the $P \times 1$ vector holding the model output (x, y, t, ω_j) for each collocation point ω_j, and the $P \times P$ matrix $\mathbf{\Phi}$ contains the P polynomials evaluated at the P collocation points. It is worth mentioning that the matrix $\mathbf{\Phi}(\omega)$ is space- and time-independent and can be produced once for the given expansion degree and parameter number, whereas the $V_c(x, y, t)$ vectors and $V_Y(x, y, t)$ are space- and time-dependent.

The efficiency of the projection method strongly rests on the selection of collocation points (Li and Zhang, 2007). Villadsen and Michelsen (1978) assert that the best selection of collocation points agrees with the roots of the polynomial of one degree higher than the order considered in the chaos expansion. This corresponds to the idea behind many numerical quadrature rules, such as Gaussian integration (Villadsen and Michelsen, 1978). Thus, the resulting polynomial surface is precise at the collocation points and applies a polynomial interpolation or extrapolation between and outside of the parameter range covered by the collocation points, respectively.

3.3   Iterative Bayesian updating and selection of reduced models

Once a response surface has been created, we can perform an approximate version of parameter updating (Eq. (7)). In this study, we apply Bayesian parameter updating (Eq. (3)) in a Monte Carlo framework (see e.g., Smith and Gelfand, 1992) applied to the response surface $\tilde{Y}_k(\omega)$. We weight N_{MC} realizations from the prior distribution $p(\omega|\widetilde{M}_k)$ of uncertain parameters with their corresponding likelihood to obtain statistics of the posterior distribution $p(\omega|\widetilde{M}_k, y_o)$. This step involves evaluation of the likelihood as in Eq. (4), but using the reduced model $\tilde{Y}_k$ instead of the original one. However, when the prior information is far from reality, it is likely that the response surface is highly imprecise and produces wrong outcomes in the parameter ranges that show high likelihood values (see Oladyshkin et al., 2013a). To avoid this, we use an advanced iterative approach for aPC-based Bayesian updating. After a first iteration of parameter updating, the realization of ω with the highest likelihood is chosen as additional collocation point and the original model (the hydro-morphodynamic model in TELEMAC-2D/SISYPHE) is executed. The result of this run is used to improve the response surface by repeated application of PCM. Once the response surface captures sufficiently well the main

features of the underlying system and the posterior distributions have stabilized during the iterative procedure (see Oladyshkin et al., 2013a), we use it for a final step of parameter updating.

Subsequently, we apply the BMS approach described in Section 3.1 to the reduced models $\tilde{M}_k$. Eq. (6) is evaluated with respect to the predictions $\tilde{Y}_k$ of the reduced models $\tilde{M}_k$ instead of the full-complexity models' predictions Y_k. As such, we perform model selection in the reduced model space and implicitly assume that these results are a good proxy for model selection in full-complexity model space. In terms of Eq. (6), we approximate $p(y_o|M_k)$ with $p(\mathbf{y}_o|\tilde{M}_k)$:

$$p(\mathbf{y}_o|M_k) \approx p(\mathbf{y}_o|\tilde{M}_k) = \int_\Omega p(\mathbf{y}_o|\tilde{M}_k, \boldsymbol{\omega})p(\boldsymbol{\omega}|\tilde{M}_k)d\boldsymbol{\omega} = \text{BME}_{\text{RM}}. \qquad (9)$$

Since the number of full-complexity model runs is limited by the extremely computational demanding simulation, we consider response surfaces of first order for BMS. Higher-order expansions would strongly increase the number of required full-complexity model runs. The computation time should be allocated considering that the response surfaces have to be constructed for each competing model. Once the best model has been identified, higher-order response surfaces can be used for refined parameter updating as discussed in Oladyshkin et al. (2013b).

3.4 Correcting Bayesian selection of reduced models

Reduced models are only a (possibly rough) approximation of the full-complexity model. Hence, any conclusions drawn from BME values based on reduced models are only valid to the degree that we believe in the accuracy of the approximation. One may accept this limitation as a "necessary evil" of BMS via model reduction, i.e. a trade-off with efficiency. Indeed, as stated earlier, BMS only becomes feasible for applications like the one presented in this study through model reduction.

Still, care must be taken when interpreting model ranking results based on model evidence for reduced models, BME_{RM}. If, using BME_{RM}, a model were penalized because its reduced variant shows a low approximation quality, model ranking via reduced models could be viewed as a conservative guess about the appropriateness of the full-complexity models. However, this is not necessarily the case: the reduced model could meet (just by chance) the observed data even better than the full-complexity model, and in that case the evidence in favor of the model would be overestimated. Hence, model ranking could turn out differently for the reduced models than for the full-complexity models, which would trigger wrong conclusions and finally possibly wrong management decisions.

To avoid such misleading results, we go one step further and propose to explicitly account for the quality of approximation by the reduced model in model ranking. We thereby hope to obtain model ranking results that are more representative of the desired model ranking results for the full-complexity models. To do so, we consider that the full-complexity model M_k cannot be fully represented by the reduced model $\tilde{M}_k$ and contains an approximation error E_k, i.e. $M_k = \tilde{M}_k + E_k$. Employing this property of additivity, Eq. (6) can be reformulated as follows:

$$p(\mathbf{y}_o | M_k) = \int_\Omega p(\mathbf{y}_o | \tilde{M}_k + E_k, \boldsymbol{\omega}) p(\boldsymbol{\omega} | M_k) d\boldsymbol{\omega}. \tag{10}$$

Considering the independence of the approximation error from the measurement errors that define the likelihood function $p(y_o|M_{k)}$ in Eq. (9), Eq. (10) can be rearranged as:

$$p(\mathbf{y}_o | M_k) = \int_\Omega p(\mathbf{y}_o | \tilde{M}_k, \boldsymbol{\omega}) p(M_k | \tilde{M}_k, \boldsymbol{\omega}) p(\boldsymbol{\omega} | M_k) d\boldsymbol{\omega}. \tag{11}$$

Here, $p(M_k | \tilde{M}_k, \boldsymbol{\omega})$ denotes the likelihood of the predictions made by the full-complexity model, Y_k, given the approximative predictions of the reduced model $\tilde{Y}_k$:

$$p(M_k | \tilde{M}_k, \boldsymbol{\omega}) = (2\pi)^{-N_s/2} |\mathbf{S}|^{-\frac{1}{2}}$$
$$\cdot \exp\left[-\frac{1}{2}(\mathbf{Y}_k(\boldsymbol{\omega}) - \tilde{\mathbf{Y}}_k(\boldsymbol{\omega}))^T \mathbf{S}^{-1}(\mathbf{Y}_k(\boldsymbol{\omega}) - \tilde{\mathbf{Y}}_k(\boldsymbol{\omega}))\right], \tag{12}$$

where $Y_k(\boldsymbol{\omega})$ is the full-complexity model run for the parameter set $\boldsymbol{\omega}$, $\tilde{Y}_k(\boldsymbol{\omega})$ is the corresponding response surface and $\mathbf{S}$ is the covariance matrix of approximation errors, i.e. $\mathbf{S}$ reflects the uncertainty in the approximation.

Without loss of generality, we can multiply and divide the right hand side of Eq. (11) by the BME value for the reduced model, $p(\mathbf{y}_o|\tilde{M}_k)$, obtained from Eq. (9):

$$p(\mathbf{y}_o | M_k) = p(\mathbf{y}_o | \tilde{M}_k) \int_\Omega p(M_k | \tilde{M}_k, \boldsymbol{\omega}) \frac{p(\mathbf{y}_o | \tilde{M}_k, \boldsymbol{\omega}) p(\boldsymbol{\omega} | M_k)}{p(\mathbf{y}_o | \tilde{M}_k)} d\boldsymbol{\omega}. \tag{13}$$

Employing Bayes' Theorem (Eq. (3)) and taking into consideration that the parameter prior is the same for the full model M_k and its reduced model $\tilde{M}_k$, $(p(\boldsymbol{\omega}|M_k) = p(\boldsymbol{\omega}|\tilde{M}_k))$, we obtain the following representation of the full-model BME:

$$p(\mathbf{y}_o | M_k) = p(\mathbf{y}_o | \tilde{M}_k) \int_\Omega p(M_k | \tilde{M}_k, \boldsymbol{\omega}) p(\boldsymbol{\omega} | \tilde{M}_k, \mathbf{y}_o) d\boldsymbol{\omega}. \tag{14}$$

Eq. (14) explicitly demonstrates that the full-model BME value (BME$_\text{corr}$) is a product of the BME value obtained for the reduced model (BME$_\text{RM}$) and the correction factor (Weight$_\text{RM}$) reflecting the credibility of the reduced model as representative of the full-complexity model:

$$\text{BME}_\text{corr} = \text{BME}_\text{RM} \cdot \text{Weight}_\text{RM}, \tag{15}$$

where each term of Eq. (15) is defined accordingly:

$$\text{BME}_{\text{corr}} = p(\mathbf{y}_o|M_k), \tag{16}$$

$$\text{BME}_{\text{RM}} = p(\mathbf{y}_o|\tilde{M}_k), \tag{17}$$

$$\text{Weight}_{\text{RM}} = \int_{\Omega} p(M_k|\tilde{M}_k, \boldsymbol{\omega}) p(\boldsymbol{\omega}|\tilde{M}_k, \mathbf{y}_o) d\boldsymbol{\omega}. \tag{18}$$

Hence, the credibility factor Weight$_{\text{RM}}$ is defined as the integral of the approximation likelihood $p(\text{M}_k|\tilde{M}_k, \boldsymbol{\omega})$ over the posterior parameter space.

From a practical point of view, the correction factor defined in Eq. (15) can be estimated from the available collocation points $\boldsymbol{\omega}$:

$$\text{Weight}_{\text{RM}} \approx \sum_{i=1}^{P} p(M_k|\tilde{M}_k, \boldsymbol{\omega}_i) \cdot p(\boldsymbol{\omega}_i|\tilde{M}_k, \mathbf{y}_o). \tag{19}$$

This yields a corrected BME value that is more representative (although still an approximation) of the full-complexity model's BME. We will compare the rankings based on Eqs. (15) (BME$_{\text{corr}}$) and (9) (BME$_{\text{RM}}$) and discuss the impact of approximation accuracy on model ranking results in Section 4.3.

Additionally, we would like to remark that the correction factor in Eq. (18) could be approximated using only the maximum a posteriori parameter set $\boldsymbol{\omega}_*$ obtained from Bayesian updating. This very rough estimation could be useful as a first glimpse at approximation quality. Then, evaluating Eq. (19) for the maximum a posteriori parameter set, $\boldsymbol{\omega}_*$, obtained from Bayesian updating of the reduced model would simplify to:

$$\begin{aligned}
\text{Weight}_{\text{RM}} &\approx (2\pi)^{-N_s/2}|\mathbf{S}|^{-\frac{1}{2}} \cdot \exp\left[-\frac{1}{2}\mathbf{F}_*^T\mathbf{S}^{-1}\mathbf{F}_*\right] \\
&= (2\pi)^{-N_s/2}\frac{1}{\prod_j |F_{*j}|} \cdot \exp\left[-\frac{1}{2}N_s\right] \\
&= \frac{1}{\sqrt{(2\pi e)^{N_s}}}\frac{1}{\prod_j |F_{*j}|},
\end{aligned}$$

with $j = 1 \ldots N_s$ and $\mathbf{F}_*$ corresponding to the vector of approximation errors obtained for the maximum a posteriori parameter set $\boldsymbol{\omega}_*$: $\mathbf{F}_* = \mathbf{Y}_k(\boldsymbol{\omega}_*) - \tilde{Y}_k(\boldsymbol{\omega}_*)$. Assuming that the approximation errors are independent and Gaussian distributed, their covariance matrix turns into a diagonal matrix $\mathbf{S} = (diag[\mathbf{F}_*])^2$. The estimation based on the maximum a posteriori estimate is very approximate; only if the posterior was (very close to) a Dirac function, this version of the correction factor would be reliable.

3.5   Scenarios for model selection

We will consider two scenarios of the lower Rhine case study for BMS. The first test case is a synthetic experiment dedicated to the proof of the presented Bayesian concept for model selection (Section 3.5.1). The second test case focuses on the transfer of the established framework to a real-data application (Section 3.5.2).

For each test case, we will perform Bayesian updating in a brute force Monte Carlo framework (Hammersley, 1960), using an ensemble of 20,000 realizations on the reduced model. It is the most direct and straightforward numerical implementation of Bayes theorem, which can consider arbitrary non-linear model equations and arbitrary distribution forms. Each realization corresponds to a randomly chosen set of parameter values, drawn from the prescribed uniform parameter distributions (Section 3.1) with bounds given in Table 1. It can be seen as numerical estimation of the reduced model-based integral in Eq. (9). In the current study, sampling error of the Monte Carlo technique is negligible in comparison to the approximation error in the reduced models.

Bed evolution predictions based on these parameter combinations are then compared to the observation data in order to obtain a posterior, updated distribution of parameters and predictions. After iterative improvement of the reduced model during the Bayesian updating procedure, we will estimate BME for each model based on the observation data set for bed evolution as described in Section 3.3.

After iterative improvement of the reduced model during the Bayesian updating procedure, we will estimate BME for each model based on the observation data set for bed evolution as described in Section 3.3. Note that sampling from the prior has been shown to be biasfree and most reliable when determining BME (e.g., Schöniger et al., 2014; Zeng et al., 2018). The only disadvantage of this bias-free method is that it needs a large number of samples; however, with the negligible run time of the reduced models (milliseconds), there is practically no limitation in sample size anymore. We therefore choose the most expensive but also most reliable method for evidence estimation over more efficient but bias-prone methods such as, e.g., MCMC with the harmonic mean estimator (Newton and Raftery, 1994).

3.5.1 Synthetic benchmark experiment

The goal of the experiment is to verify that the BMS framework is able to infer the true morphodynamic model out of the six sediment transport models listed in Section 2, under the ideal condition that the true model is actually part of the model set. This is of course never the case in reality, but allows us to verify that under ideal conditions, the proposed framework fulfills its purpose. We choose an arbitrary parameter set for the Rhine model from the parameter ranges in Table 1 to generate 111 sample points of synthetic measurements of bed evolution. In this case, the Wu formula was chosen and the simulated bed evolution after 17 days was taken as synthetic measurements.

We assume a relatively low level of measurement error for the synthetic bed evolution data with a standard deviation of 0.03 m. This assumption enables us to find out whether BMS is able to identify

the true model under relatively well-posed conditions (i.e., measurements are not overly noisy). Moreover, ten iterations are allowed to improve the
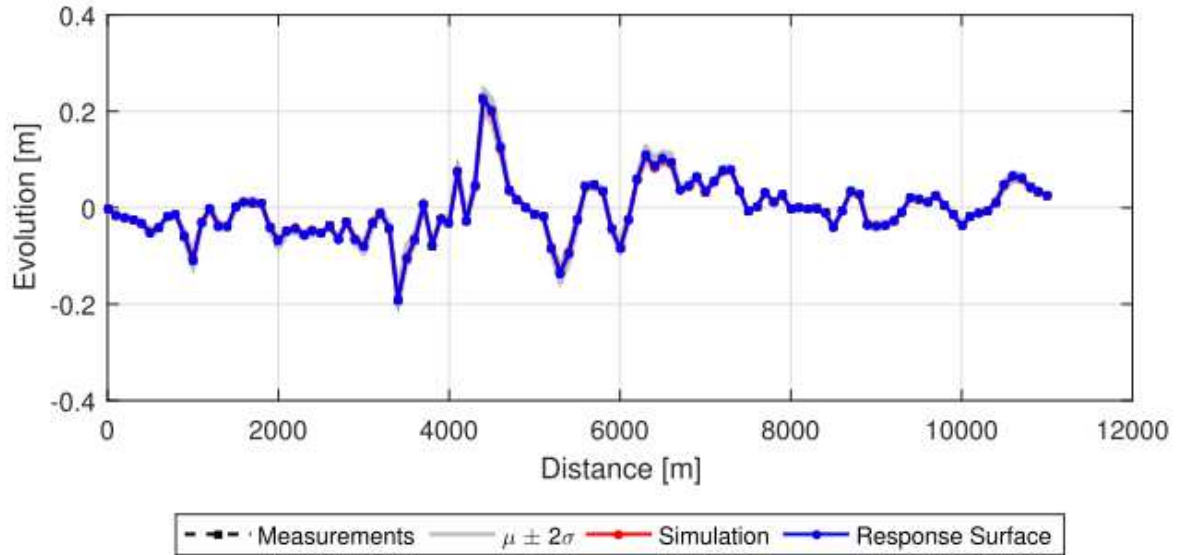

Fig. 2: *Bayesian updating of Wu model: synthetic benchmark experiment.*

accuracy of the reduced models in reproducing the predictions of the full-complexity models.

3.5.2 Real-data test case

The real-data test case focuses on the transfer of the synthetically verified model selection framework to a real-data application. To this end, cross-profile soundings of the river bottom collected by the German Federal Waterways and Shipping Administration (WSV) in 2004 are used. 111 sample points of the averaged bottom levels, as explained in Section 2.2, are considered as measurements for model selection.

Four most promising competing sediment transport models for the current case study have been considered, namely (Hunziker, 1995; Meyer-Peter and Müller, 1948; van Rijn, 1984; Wu et al., 2000).

A measurement error standard deviation of 0.05 is assumed for this test case. Although this error is a little larger than assumed in the synthetic test case, the much longer time series nevertheless assures a strong updating effect and results in a low posterior parameter and predictive uncertainty (see Section 4.2). Due to a very strong computational time demands of the full-complexity models only five iterations are allowed to improve the accuracy of the reduced models.

4 Bayesian selection of sediment transport models

4.1 Results of synthetic benchmark experiment

The bed evolution after 17 days predicted by the hydro-morphodynamic model based on the Wu sediment transport model is illustrated in Fig. 2. Here and in the following figures, the horizontal axis shows the distance from the inlet boundary (km 738.5 of the lower Rhine), whereas the vertical axis represents the bed evolution. These predictions have been used as synthetic measurement data to rank the set of competing transport models. The synthetic measurements (black dashed line) simulated by the full-complexity model (red line) and the simulation with response surface (blue line) overlap entirely. This indicates that the model reduction technique embedded in the Bayesian selection framework is successful in representing the full-complexity model in the relevant parameter region after Bayesian updating with ten iterations. Note that both the full-complexity model simulation and the response surface are shown here for the parameter realization with the highest posterior probability after Bayesian updating. The remaining parameter uncertainty after Bayesian updating results in predictive uncertainty, which is represented by the gray-shaded credible interval of ±2 standard deviations.

Out of the five alternative model formulations, we exemplarily show the output of the Bayesian updating procedure for the widely used Meyer-Peter and Müller formula in Fig. 3. Fig. 3 indicates that the response surface constructed during Bayesian updating is again a suitable representation of the full-complexity numerical model. The outputs for the remaining competing models can be found in Mohammadi (2017).

The BME values of the six reduced sediment transport models are listed in Table 2. As expected, the true model (here: Wu) obtains
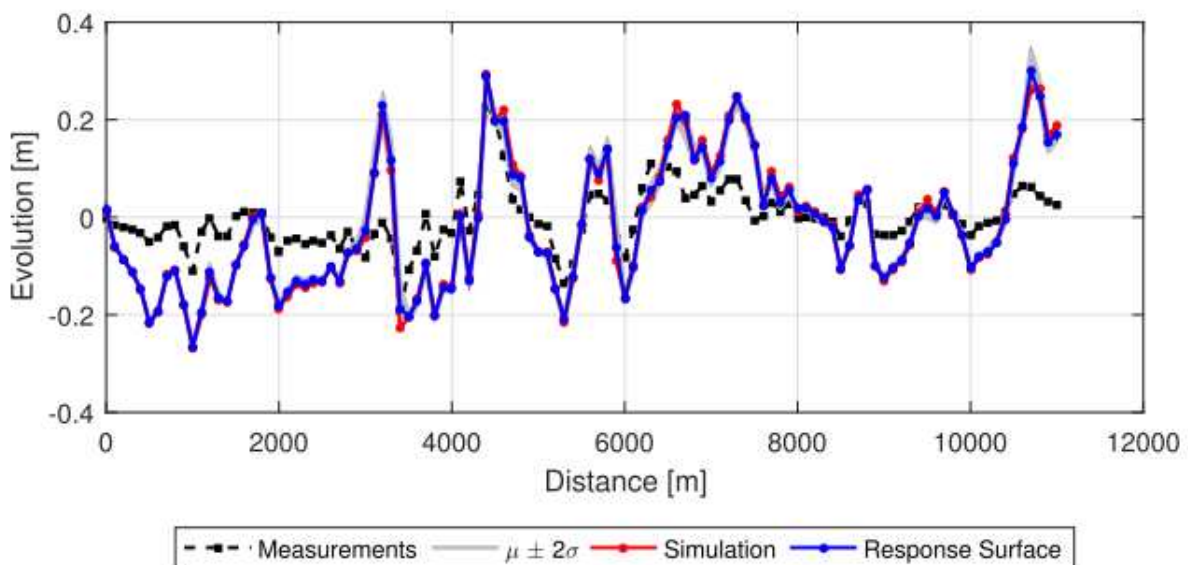

Fig. 3:   *Bayesian updating of Meyer-Peter and Müller model: synthetic benchmark experiment.*

Table 2:   *BME ranking of sediment transport models based on reduced models (synthetic benchmarking experiment).*

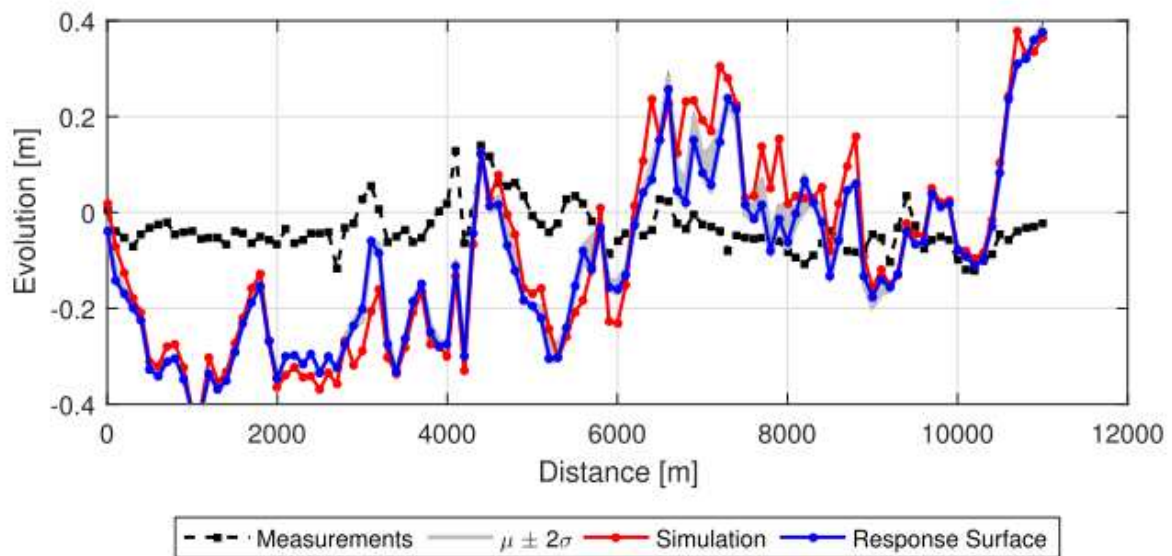| Model	Rank	BME_{RS}
Wu	1	1.14E+124
Hunziker	2	1.69E+87
Van Rijn	3	7.75E+75
Einstein-Brown	4	1.71E+37
Meyer-Peter and Müller	5	6.44E-2
Modified Engelund-Hansen	6	9.29E-83

Fig. 4:   *Bayesian updating of Meyer-Peter and Müller model: real-data rest case.*

by far the highest evidence value. The Bayes factor with respect to the second-best model (here: Hunziker) is on the order of E+36, which represents more than decisive evidence on the scale proposed by Jeffreys (1967) (see Section 3.1). The ranking suggested by the BME values proves that the proposed BMS framework is able to correctly and confidently identify the true data-generating model if it is in the set. This result enables us to proceed with the real-data test case where no true model exists; instead we are interested in identifying the most suitable model in the set in the light of the actually observed data.

4.2 Results of real-data test case

In this Section, we present the results of the Bayesian analysis using real data and four pre-selected competing sediment transport models. Figs. 4–7 show the measured bed evolution, the original model's simulation output and the response surface using the most probable parameter set for all four competing formulas. All models show too large magnitudes of bed evolution, especially the latest, which could be due to uncertain boundary conditions. The obvious conclusion is that none of the models is overly good in matching the real data, which is partly due to the practical constraints that

limit the available simulation time for calibration purposes. Since all models are suffering from this issue, it is not a trivial task to choose the single best one. Therefore, we apply the proposed Bayesian framework based on reduced models to identify
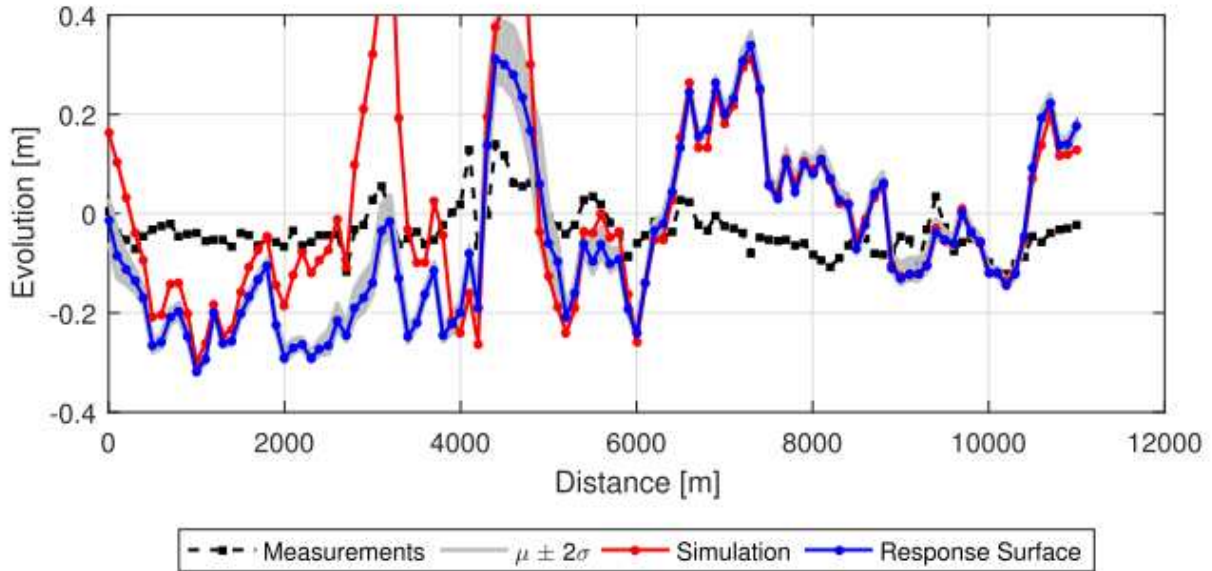

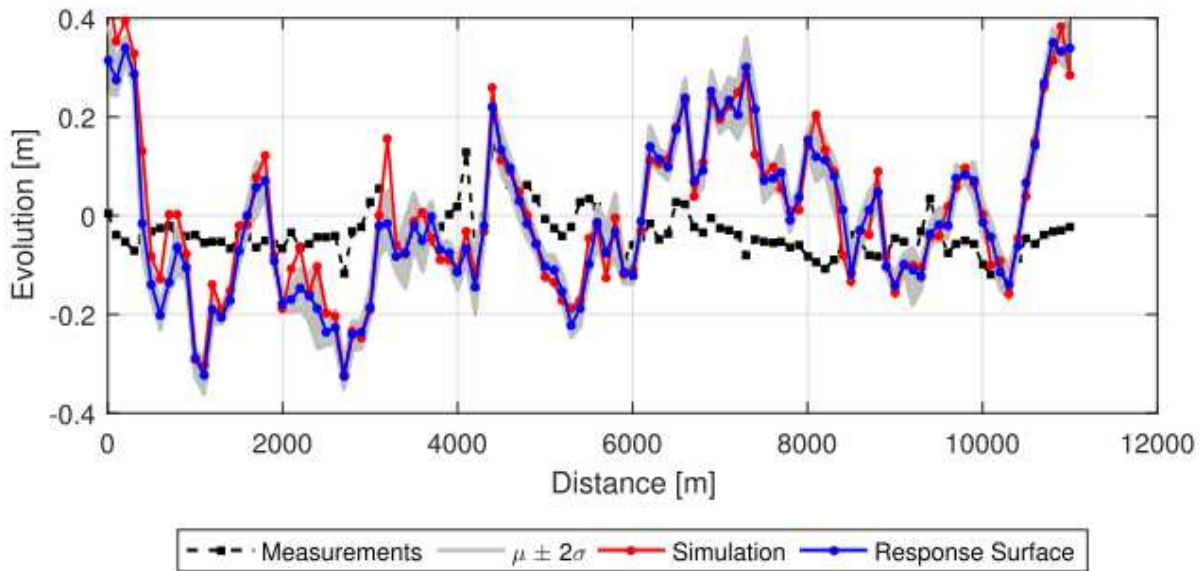*Fig. 5:*   *Bayesian updating of van Rijn model: real-data rest case.*

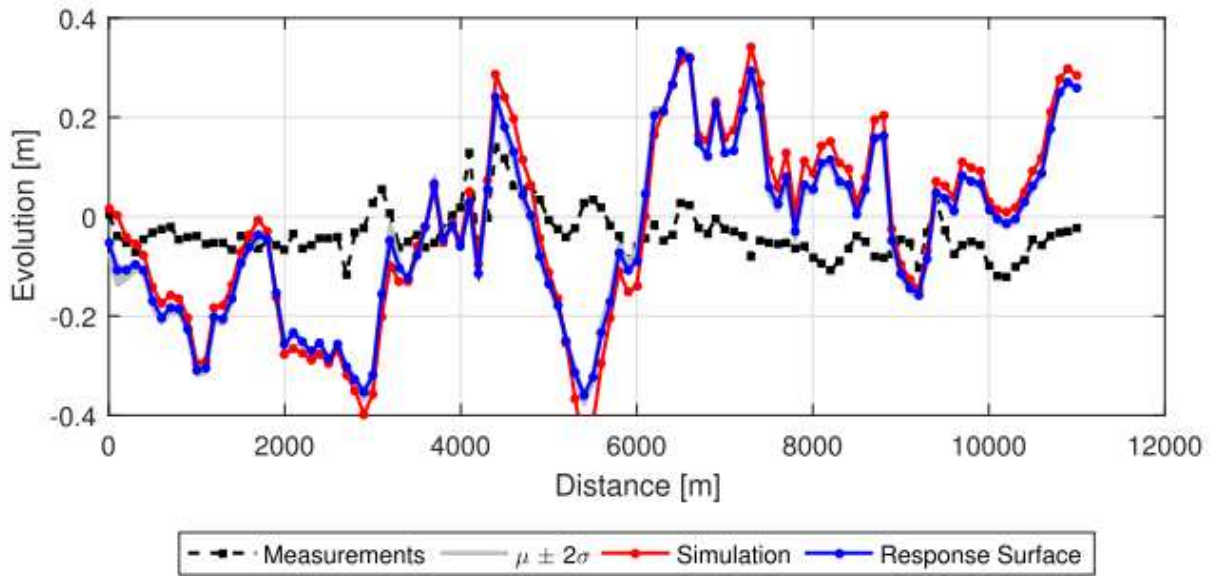*Fig. 6:*   *Bayesian updating of Hunziker model: real-data rest case.*

Fig. 7:   *Bayesian updating of Wu model: real-data rest case.*

the most suitable model under the given strong computational time constraints.

The plots in Fig. 4 illustrate the simulation output of the hydro-morphodynamic model including the widely used Meyer-Peter and Müller formula. As it is visible from Fig. 4, the updated response surface after five iterations could capture the pattern of the original model in TELEMAC-2D-SISYPHE, but was not utterly successful in representing its magnitude. However, the misfit between the full-complexity numerical model and its response surface (Fig. 4) implies that the reduced model is not a sufficient approximation of the full-complexity model. This is probably a shortcoming of the linear approximation that we chose here due to the given computational limitations presented in Section 2.2. It is to be expected that higher-order response surface approximations will have higher adaptivity and could improve the approximation during Bayesian updating.

Fig. 5 illustrates the results for the van Rijn Model. Apparently, the response surface could not capture accurately the full-complexity model behavior before the fifth kilometer from the inlet boundary, particularly the peak values. However, this misfit changes significantly towards the outlet edge. Fig. 6 shows the results for the Hunziker model. In contrast to the other reduced models, here the credible intervals (gray intervals) of the final response surface after five iterations contain most of the bed evolution predictions made by the full-complexity model.

Table 3:   *BME ranking of sediment transport models based on reduced models (real-data test case).*

Model	Rank	BME_{RS}
Van Rijn	1	3.73E+36
Hunziker	2	1.44E+36
Wu	3	3.79E+27
Meyer-Peter and Müller	4	1.84E+10

Table 4:   Corrected BME ranking of sediment transport models based on all available collocation points (synthetic benchmarking experiment).

Model	Rank	BME_{RM}	$Weight_{RM}$	BME_{corr}
Wu	1	1.14E+124	1.48E-71	1.69E+53
Hunziker	2	1.69E+87	2.07E-70	3.51E+17
Van Rijn	3	7.75E+75	3.49E-72	2.70E+4
Einstein-Brown	4	1.71E+37	1.35E-75	2.31E-38
Meyer-Peter and Müller	5	6.44E-2	1.73E-74	1.12E-75
Modified Engelund-Hansen	6	9.29E-83	4.18E-119	3.88E-201

Fig. 7 shows the Bayesian updating results for the Wu Model. The response surface built in the Bayesian framework captures the overall pattern, but it slightly fails in disclosing the magnitude of the original model in TELEMAC-2D-SISYPHE. The credible intervals of the response surface (gray intervals in Fig. 7) are considerably small. The model ranking results based on BME estimation for the four competing models are listed in Table 3. The hydro-morphodynamic model based on the van Rijn sediment transport formula scores the highest BME value and is therefore ranked first. This is due to the fact that this model's response surface captures the observed data best among the competing models. With respect to the second-best model (here: Hunziker), the van Rijn model achieves a Bayes factor of 2.6, which, according to the rule of thumb by Jeffreys (1967), is not significant. Hence, the van Rijn model and the Hunziker model seem to be practically equally suitable for the case study presented here. Both models achieve a Bayes factor on the order of E+8 when compared to the third-best model (here: Wu), which suggests that the Wu model (and the Meyer-Peter and Müller model which ranks fourth) should be rejected against the two best models.

4.3   Impact of model reduction accuracy on model ranking

According to Section 4.2, the van Rijn model and the Hunziker model (as represented by their response surfaces) have been selected as suitable models based on the real bed evolution measurements. However, the response surface for the van Rijn model significantly deviates from the simulation output of the corresponding full-complexity model as obvious from Fig. 5. Indeed, the response surfaces of the four competing sediment transport models vary greatly in approximation quality. For example, the response surface constructed for the Hunziker model captures the full-complexity model's simulations fairly well. These findings suggest that model ranking based on reduced models is potentially unfair - bad full-complexity models could be ranked higher than they should because their response surface is by chance closer to the measurements, and vice versa. This is why we see a need to introduce a correction factor for the reduced models that shall correct the model ranking in view of the approximation quality as proposed in Section 3.4. We thereby hope to obtain a model ranking that is representative of the full-complexity models and not (only) of the reduced models constructed under computational time constraints.

We apply Eqs. (19) and (20) to assess the reduced models' quality, based on all collocation points and based only on the maximum a posteriori parameter estimate, respectively. Then, we obtain two

versions of the corrected BME value from Eq. (15). The resulting model rankings are listed in Tables 4 and 5, respectively, for the synthetic benchmark-

Table 5:   Corrected BME ranking of sediment transport models based only on maximum a posteriori parameter set (synthetic benchmarking experiment).

Model	Rank	BME_{RM}	$Weight_{RM}$	BME_{corr}
Wu	1	1.14E+124	9.42E-66	1.07E+59
Hunziker	2	1.69E+87	2.68E-67	4.54E+20
Van Rijn	3	7.75E+75	8.32E-69	6.45E+7
Einstein-Brown	4	1.71E+37	2.99E-68	5.09E-31
Meyer-Peter and Müller	5	6.44E-2	3.45E-67	2.22E-68
Modified Engelund-Hansen	6	9.29E-83	3.51E-69	3.27E-151

Table 6:   Corrected BME ranking of sediment transport models based on all available collocation points (real-data test case).

Model	Rank	BME_{RM}	$Weight_{RM}$	BME_{corr}
Van Rijn	1	3.73E+36	1.07E-72	3.99E-36
Hunziker	2	1.44E+36	3.18E-73	4.57E-37
Wu	3	3.79E+27	8.05E-72	3.05E-44
Meyer-Peter and Müller	4	1.84E+10	5.16E-98	9.49E-88

Table 7:   Corrected BME ranking of sediment transport models based only on maximum a posteriori parameter set (real-data test case).

Model	Rank	BME_{RM}	$Weight_{RM}$	BME_{corr}
Hunziker	1	1.44E+36	1.79E-68	2.56E-32
Van Rijn	2	3.73E+36	1.94E-69	7.24E-33
Wu	3	3.79E+27	2.67E-68	1.02E-40
Meyer-Peter and Müller	4	1.84E+10	1.06E-68	1.94E-58

ing experiment and in Tables 6 and 7, respectively, for the real-data test case.

For the synthetic benchmark experiment, the approximation quality of the reduced models does not have an impact on the final model ranking since the differences in matching the synthetic data are much larger than the differences in approximation quality. In such a case, the approximation error is small compared to the error in fitting the data, and hence a correction (reweighting) of the BME values obtained from reduced models would not be necessary. However, the effort to determine the correction factor is negligible since all required model results are obtained during the model selection analysis anyway and hence we recommend to perform the correction to be on the safe side. Future applications of the correction factor in various synthetic setups will provide further insight into its potential benefit under specific conditions.

For the real-data test case, the ranking of the two best models is indeed sensitive to the applied penalty for approximation error. The reduced model variant of the van Rijn model scores the highest BME value (BME_{RM}), but shows partially huge deviations from its full-complexity variant (reflected

by a low Weight$_{RM}$) and is, therefore, down-graded in its ranking according to the corrected value BME$_{corr}$ based on the maximum a posteriori parameter set. The second-best model based on BME$_{RM}$, the Hunziker model, then turns out as the winner because it obtains the highest likelihood in terms of agreement with its full-complexity variant and therefore scores the highest corrected BME value. However, when taking into account all available collocation points, the Van Rijn model receives a slightly weaker penalty than the Hunziker model and therefore remains to be ranked first. In case of the Wu and the Meyer-Peter and Müller models, both implementations of the correction factor only lead to a rescaling of the BME value but do not reverse the ranking. A Bayes factor on the order of E+7 (according to both error estimates) and larger still suggests to confidently reject the third- and fourth-ranking models against the two best models in the set. Between these two best models, the correction factor based on all collocation points has revealed a Bayes factor of 8.7 in favor of the Van Rijn model, which now translates into substantial evidence in favor of this model. The contradiction between the correction factors based on all available collocation points and based on only the maximum a posteriori parameter estimate is in line with the fact, that for reliable BME estimation, a good representation of the full parameter space is important. To obtain even more realistic correction factor values, alternative approaches to approximate the integral in Eq. (14) could be drawn upon; this is however beyond the scope of this study.

Thus, we find that the proposed correction for BME values based on reduced models helps to objectively penalize inaccurate reduced models. As we had feared, the inaccuracy of reduced models does not automatically lead to low evidence values but needs to be corrected for, otherwise the evidence of a lucky reduced model may be overestimated. In this specific case, the confidence in model selection was increased through applying the correction based on all available collocation points.

5   Summary & conclusions

In this paper, we have addressed Bayesian selection of hydro-morphodynamic models under strong computational time constraints. We incorporate BME as a model evaluation yardstick for ranking competing models that represent the hydro-morphodynamics. It accounts for measurement errors in the observation data and it also takes uncertainty of modeling parameters into consideration. Since we are limited to a maximum of approximately 10 runs per model in our case study, we cannot use popular sampling strategies that aim at increased efficiency of numerical integration to obtain Bayesian model weights. Instead, we make use of the concept of model reduction that is commonly employed to be able to perform stochastic calibration and uncertainty quantification for these types of models. Hence, we render the BMS framework applicable for computationally intensive hydro-morphodynamic models, by combining it with model reduction based on arbitrary polynomial chaos expansion. The reduced model represented by the response surface captures the hydro-morphodynamic model's dependence on uncertain parameters. Additionally, to better adapt the response surface to the original full-complexity model and to avoid misleading inferences of parameters, we iteratively update the response surface in the best-fit parameter range against the observation data. To do so, we increase the precision of the response surface around the current iteration of the posterior distribution by adding the current maximum probability parameter set into

the collection of model runs that is used to build the response surface. Moreover, taking into consideration that the BME value based on the reduced model is merely an estimate of the true BME value, we newly introduce a correction factor that corrects the model ranking in view of the approximation quality. The calculation of the correction factor makes use of the model runs needed for establishing the reduced models, and hence avoids additional computational effort. The proposed correction factor is a first step towards obtaining reliable model ranking results for reduced models in practical applications, since a comparison with a reference solution for model weights is typically infeasible, given that a reliable estimation of model weights requires on the order of millions of full-complexity model runs.

To assess the newly developed framework, we have established a benchmark study based on a river model for the lower Rhine, where synthetic bed evolution measurements are generated from one of the competing models. The BMS framework leads us to the correct model choice, which is the model used to generate the artificial bed evolution observations. To transfer the presented BMS framework to large-scale problems, we have simulated the lower Rhine model for 2 years and matched it against real measurement data of bed evolution. The hydro-morphodynamic model embedding the van Rijn sediment transport formula is ranked first. However, the response surfaces for all considered models cannot fully capture the corresponding full-complexity model and cannot be seen as fully reliable. With the proposed correction factor based on all available collocation points, model selection becomes more confident and identifies the van Rijn model as substantially more suitable than the Hunziker model. This model choice result is specific to the featured river segment and data, and should not be seen as a general recommendation. In our case, differences in model complexity were small compared to the vast differences in model performance. Hence, we benefited from the integral nature of performance assessment via BME, while in other applications additionally the penalty for overly complex models may be a decisive reason to choose the BMS framework for model selection.

We conclude that blind use of Bayesian principles under strong computational time constraints could be misleading and an assessment of approximation quality of reduced models should be taken into consideration. Overall, the presented reduced model-based Bayesian selection framework is helpful to objectively select between different available models according to their capability to capture the observation data, even if the models are computationally too costly for standard procedures of BMS. These properties of our framework make it applicable for a wide range of models, such as subsurface flow models, reservoir models, reactive transport models or hydrogeophysical models, to name an arbitrary selection. Indeed, the proposed framework could prove useful in any modeling endeavor that faces model choice uncertainty under extreme computational time constraints, with the limitation that the dimensionality of the parameter space and the complexity of the model allows for a reasonable approximation through PCE. Moreover, the suggested concept of the correction factor could be used in combination with other model reduction techniques (Köppel et al., 2018) available in various disciplines such as uncertainty quantification, applied mathematics, sparse grid or machine learning.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2018.05.007.

References

Ackers, P., White, W.R., 1973. Sediment transport: new approach and analysis. J. Hydraul. Div. 99 (hy11), 2041–2060.

Backhaus, L., Riesterer, J., Brudy-Zippelius, T., Wenka, T., 2014. Comparison of morphological predictions in the Lower Rhine River by means of a 2-D and 3-D model and in situ measurements. In: Proceedings of River Flow 2014, Lausanne.

Batalla, R.J., 1997. Evaluating bed-material transport equations using field measurements in a sandy gravel-bed stream, Arbúcies River, NE Spain. Earth Surf. Process. Landf. 22 (2), 121–130.

Bazargan, H., Christie, M., Elsheikh, A.H., Ahmadi, M., 2015. Surrogate accelerated sampling of reservoir models with complex structures using sparse polynomial chaos expansion. Adv. Water Resour. 86, 385–399.

Beckers, F., Noack, M., Wieprecht, S., 2017. Uncertainty analysis of a 2D sediment transport model: an example of the Lower River Salzach. J. Soils Sedim. 1–12.

Brunetti, C., Linde, N., Vrugt, J.A., 2017. Bayesian model selection in hydrogeophysics: application to conceptual subsurface models of the South Oyster Bacterial Transport Site, Virginia, USA. Adv. Water Resour. 102, 127–141.

Camacho, R.A., Martin, J.L., McAnally, W., Díaz-Ramirez, J., Rodriguez, H., Sucsy, P., Zhang, S., 2015. A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling. JAWRA J. Am. Water Resour.Assoc. 51 (5), 1372–1393.

Chollet, J.P., Cunge, J.A., 1980. New interpretation of some head loss flow velocity relationship for deformable bed. J. Hydraul. Res. 17 (1), 1–13.

Cole, C.V., Tarapore, Z.S., Kanhere, V.N., Dixit, J.G., 1973. Sediment discharge formulae- a comparative analysis. Sedim. Transp. 1, 547–560.

Draper, D., 1995. Assessment and propagation of model uncertainty. J. R. Stat. Soc. Ser. B 57, 45–97.

Einstein, H.A., et al., 1950. The Bed-Load Function for Sediment Transportation in Open Channel Flows. (Vol. 1026).. Washington DC: US Department of Agriculture.

Elsheikh, A.H., Jackson, M.D., Laforce, T.C., 2012. Bayesian reservoir history matching considering model and parameter uncertainties. Math. Geosci. 44 (5), 515–543.

Engelund, F., 1974. Flow and bed topography in channel bends. J. Hydraul. Div. 100 (11), 1631–1648.

Engelund, F., Hansen, E., 1967. A Monograph on Sediment Transport in Alluvial Streams. Technical Report. Technical University of Denmark 0stervoldgade 10, Copenhagen K.

Gilbert, G.K., Murphy, E.C., 1914. The Transportation of Debris by Running Water. No. 86). US Government Printing Office.

Gomez, B., Church, M., 1989. An assessment of bed load sediment transport formulae for gravel bed rivers. Water Resour. Res. 25 (6), 1161–1186.

Gull, S.F., 1988. Bayesian inductive inference and maximum entropy. In: Maximum-Entropy and Bayesian Methods in Science and Engineering. Springer, Dordrecht, pp. 53–74.

Habibi, M., 1994. Sediment Transport Estimation Methods in River Systems. University of Wollongong Ph.D. thesis,.

Hammersley, J.M.M., 1960. Monte Carlo methods for solving multivariable problems. Ann. N.Y. Acad. Sci. 86 (3), 844–874.

Hervouet, J.-M., 2000. Telemac modelling system: an overview. Hydrol. Process. 14 (13), 2209–2210.

Hervouet, J.-M., 2007. Hydrodynamics of Free Surface Flows: Modelling with the Finite Element Method. John Wiley & Sons.

Hubbell, D.W., Matejka, D.Q., 1959. Investigations of Sediment Transportation, Middle Loup River at Dunning, Nebraska: With Application of Data from Turbulence Flume. (No. 1476).. Geological Survey (US).

Hunziker, R.P., 1995. Fraktionsweiser Geschiebetransport. ETH Zurich Ph.D. thesis.

Jeffreys, S.H., 1967. Theory of Probability: 3d Ed. Clarendon Press.

Jia, W., McPherson, B., Pan, F., Dai, Z., Xiao, T., 2018. Uncertainty quantification of CO_2 storage using Bayesian model averaging and polynomial chaos expansion. Int. J. Greenh. Gas Control 71, 104–115.

Jin, B., 2008. Fast Bayesian approach for parameter estimation. Int. J. Numer. Methods Eng. 76 (2), 230–252.

Johnson, J.W., 1943. Laboratory Investigations on Bed-load Transportation and Bed Roughness. U.S. Department of Agriculture, Soil Conservation Service.

Jordan, P.R., 1965. Fluvial Sediment of the Mississippi river at St. Louis, Missouri. US Government Printing Office.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795.

Koch, F.G., Flokstra, C., 1980. Bed Level Computations for Curved Alluvial Channels: Prepared for the 19th IAHR Congress, New Delhi, India, February 1981. Waterloopkundig Laboratorium.

Köppel, M., Franzelin, F., Kröker, I., Oladyshkin, S., Santin, G., Wittwar, D., Barth, A., Haasdonk, B., Nowak, W., Pflüger, D., Rohde, C., 2018. Comparison of data-driven uncertainty quantification methods for a carbon dioxide storage benchmark scenario. arXiv:1802.03064.

Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. Water Resour. Res. 49 (5), 2664–2682.

Li, H., Zhang, D., 2007. Probabilistic collocation method for flow in porous media: comparisons with other stochastic methods. Water Resour. Res. 43 (9), 1–13.

Li, J., Marzouk, Y.M., 2014. Adaptive construction of surrogates for the Bayesian solution of inverse problems. SIAM J. Scient. Comput. 36 (3), A1163–A1186.

Liang, M., Voller, V.R., Paola, C., 2015. A reduced-complexity model for river delta formation–Part 1: modeling deltas with channel dynamics. Earth Surf. Dyn. 3 (1), 67–86.

Liu, P., Elshall, A.S., Ye, M., Beerli, P., Zeng, X., Lu, D., Tao, Y., 2016. Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods. Water Resour. Res. 52 (2), 734–758. https://doi.org/10.1002/2014WR016718.

López, R., Vericat, D., Batalla, R.J., 2014. Evaluation of bed load transport formulae in a large regulated gravel bed river: The lower Ebro (NE Iberian Peninsula). J. Hydrol. 510, 164–181.

Lötgering-Lin, O., Schöniger, A., Nowak, W., Gross, J., 2016. Bayesian model selection helps to choose objectively between thermodynamic models: a demonstration of selecting a viscosity model based on entropy scaling. Ind. Eng. Chem. Res. 55 (38), 10191–10207.

Lu, F., Morzfeld, M., Tu, X., Chorin, A.J., 2015. Limitations of polynomial chaos expansions in the Bayesian solution of inverse problems. J. Comput. Phys. 282, 138–147.

Marshall, L., Nott, D., Sharma, A., 2005. Hydrological model selection: a Bayesian alternative. Water Resour. Res. 41 (10).

Martin, Y., Ham, D., 2005. Testing bedload transport formulae using morphologic trans- port estimates and field data: Lower Fraser River, British Columbia. Earth Surf. Process. Landf. 30 (10), 1265–1282.

Messing, S., des Wasser, A.W., Duisburg-Rhein, S., 2008. Geschiebezugabe Unterer Niederrhein. Binnenschifffahrt, Jg 63.

Meyer-Peter, E., Favre, H., Einstein, H.A., 1934. Neuere Versuchsresultate über den Geschiebetrieb. Schweizerische Bauzeitung 103 (13), 147–150.

Meyer-Peter, E., Müller, R., 1948. Formulas for bed-load transport. IAHSR 2nd Meeting, Stockholm, Appendix 2. IAHR.

Mohammadi, F., 2017. Bayesian Model Selection for Hydro-Morphodynamic Models. Master thesis. University of Stuttgart.

Müller, R., 1937. Überprüfung des Geschiebegesetzes und der Berechnungsmethode der Versuchsanstalt für Wasserbau an der ETH mit Hilfe der direkten Geschiebemessungen am Rhein. Schweizerische Bauzeitung 110 (12–15), 180–184.

Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J. R. Stat. Soc. Ser. B 56 (1), 3–48.

Nicholas, A.P., 2009. Reduced-complexity flow routing models for sinuous single-thread channels: intercomparison with a physically-based shallow-water equation model. Earth Surf. Process. Landf. 34 (5), 641–653.

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Adv. Water Resour. 34 (11), 1508–1518.

Oladyshkin, S., Class, H., Nowak, W., 2013. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: methodology and application to history matching for carbon dioxide storage in geological formations. Comput. Geosci. 17 (4), 671–687.

Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. System Saf. 106 (2010), 179–190.

Oladyshkin, S., Nowak, W., 2018. Incomplete statistical information limits the utility of high-order polynomial chaos expansions. Reliab. Eng. Syst. Saf. 169, 137–148.

Oladyshkin, S., Schröder, P., Class, H., Nowak, W., 2013. Chaos expansion based bootstrap filter to calibrate CO_2 injection models. Energy Proc. 40, 398–407.

Oliver, D.S., Reynolds, A.C., Liu, N., 2008. Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press.

Papanicolaou, A.N.T., Krallis, G., Edinger, J., 2008. Sediment transport modeling review: current and future developments. J. Hydraul. Eng. 134 (1), 1–14.

Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. Ground Water 43 (4), 597–605.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–163.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7).

Red-Horse, J.R., Benjamin, A.S., 2004. A probabilistic approach to uncertainty quantification with limited information. Reliab. Eng. Syst. Saf. 85 (1–3), 183–190.

Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Adv. Water Resour. 36, 36–50.

Riesterer, J., 2015. Bericht Validierungsfall Niederrheinmäander. Karlsruhe Institute of Technology, Institut für Wasser und Gewässerentwicklung.

van Rijn, L.C., 1984. Sediment transport, part I: bed load transport. J. Hydraul. Eng. 110 (10), 1431–1456.

Schoklitsch, A., 1934. Der Geschiebetrieb und die Geschiebefracht. Wasserkraft Wasserwirtschaft 4, 1–7.

Schöniger, A., Illman, W.A., Wöhling, T., Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. J. Hydrol. 531, 96–110.

Schöniger, A., Wöhling, T., Samaniego, L., Nowak, W., 2014. Model selection on solid ground: rigorous comparison of nine ways to evaluate Bayesian model evidence. Water Resour. Res. 50 (12), 9484–9513.

Skilling, J., et al., 2006. Nested sampling for general Bayesian computation. Bayesian Anal. 1 (4), 833–859.

Smith, A.F.M., Gelfand, A.E., 1992. Bayesian statistics without tears: a sampling–resampling perspective. Am. Stat. 46 (2), 84–88.

Talmon, A.M., Struiksma, N., Van Mierlo, M.C.L.M., 1995. Laboratory measurements of the direction of sediment transport on transverse alluvial-bed slopes. J. Hydraul. Res. 33 (4), 495–517.

Tassi, P., Villaret, C., 2014. Sisyphe v6.3 User's Manual. R&D, Électricité de France.

Villadsen, J., Michelsen, M.L., 1978. Solution of Differential Equation Models by Polynomial Approximation. Prentice-Hall.

Villaret, C., Hervouet, J.M., Kopmann, R., Merkel, U., Davies, A.G., 2013. Morphodynamic modeling using the Telemac finite-element system. Comput. Geosci. 53, 105–113.

Volpi, E., Schoups, G., Firmani, G., Vrugt, J.A., 2017. Sworn testimony of the model evidence: Gaussian mixture importance (GAME) sampling. Water Resour. Res. 53 (7), 6133–6158.

Wenka, T., Backhaus, L., Riesterer, J., 2015. Berücksichtigung dreidimensionaler Strömungseffekte und Transportphänomene in morphodynamischen Modellen von Binnenwasserstrassen. FuE-Abschlussbericht. BAW.

Wiener, N., 1938. The homogeneous chaos. Am. J. Math. 60 (4), 897–936.

Wilcock, P.R., 2001. The flow, the bed, and the transport: interaction in the flume and field. Gravel-bed Rivers V. New Zealand Hydrological Society.

Wöhling, T., Schöniger, A., Gayler, S., Nowak, W., 2015. Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction. Water Resour Res 51 (4), 2825–2846.

Wu, F.C., Chen, C.C., 2009. Bayesian updating of parameters for a sediment entrainment model via Markov Chain Monte Carlo. J. Hydraul. Eng. 135 (1), 22–37.

Wu, W., Wang, S.S.Y., Jia, Y., 2000. Nonuniform sediment transport in alluvial rivers. J. Hydraul. Resour. 38 (6), 427–434.

Wurms, S., Schröder, M., 2012. Evaluation of strategies for the acceleration of morpho- dynamic simulations against the background of waterways maintenance. In: River Flow 2012 - Int. Conf. Fluv. Hydraul.. London: Taylor & Francis, San Jose, Costa Rica, pp. 1235–1241.

Xue, L., Dai, C., Wu, Y., Wang, L., 2018. Towards improving the efficiency of Bayesian model averaging analysis for flow in porous media via the probabilistic collocation method. Water 10 (4), 412. https://doi.org/10.3390/w10040412.

Yang, C.T., Wan, S., 1991. Comparisons of selected bed-material load formulas. J. Hydraul. Eng. 117 (8), 973–989.

Zeng, X., Ye, M., Burkardt, J., Wu, J., Wang, D., Zhu, X., 2016. Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification. J. Hydrol. 535, 120–134.

Zeng, X., Ye, M., Wu, J., Wang, D., Zhu, X., 2018. Improved nested sampling and surrogateenabled comparison with other marginal likelihood estimators. Water Resour. Res. 54 (2), 797–826. https://doi.org/10.1002/2017WR020782.

Autoren

Farid Mohammadi

Rebekka Kopmann

Anneli Guthke

Sergey Oladyshkin
sergey.oladyshkin@iws.uni-stuttgart.de

Wolfgang Nowak