Received: 24 July 2021 | Revised: 14 October 2021 | Accepted: 16 October 2021

DOI: 10.1002/esp.5285

RESEARCH ARTICLE

ESPL WILEY

Fuzzy map comparisons enable objective hydro-morphodynamic model validation

Beatriz Negreiros | **Sebastian Schwindt** | **Stefan Haun** | **Silke Wieprecht**

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

Correspondence
Beatriz Negreiros, Department of Hydraulic Engineering and Water Resources Management, Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Baden Württemberg D-70569, Germany.
Email: beatriz.negreiros@iws.uni-stuttgart.de

Summary

Numerical modeling represents a state-of-the-art technique to simulate hydro-morphodynamic processes in river ecosystems. Numerical models are often validated based on observed topographic change in the form of pixel information on net erosion or deposition over a simulation period. When model validation is performed by a pixel-by-pixel comparison of exactly superimposed simulated and observed pixels, zero or negative correlation coefficients are often calculated, suggesting poor model performance. Thus, a pixel-by-pixel approach penalizes quantitative simulation errors, even if a model conceptually works well. To distinguish between reasonably well-performing and non-representative models, this study introduces and tests fuzzy map comparison methods. First, we use a fuzzy numerical map comparison to compensate for spatial offset errors in correlation analyses. Second, we add a level of fuzziness with a fuzzy kappa map comparison to additionally address quantitative inaccuracy in modeled topographic change by categorizing data. Sample datasets from a physical lab model and datasets from a 6.9 km long gravel–cobble bed river reach enable the verification of the relevance of fuzzy map comparison methods. The results indicate that a fuzzy numerical map comparison is a viable technique to compensate for model errors stemming from spatial offset. In addition, fuzzy kappa map comparisons are suitable for objectively expressing subjectively perceived correlation between two maps, provided that a small number of categories is used. The methods tested and the resulting spatially explicit comparison maps represent a significant opportunity to improve the evaluation and potential calibration of numerical models of river ecosystems in the future.

1 | INTRODUCTION

Two-dimensional (2D) numerical modeling is a state-of-the-art tool in environmental and earth sciences to simulate hydrodynamic and morphodynamic (hydro-morphodynamic) processes in river ecosystems (Ganju et al., 2016). Typically, setting up a state-of-the-art hydro-morphodynamic numerical model involves: (1) acquiring input data of the terrain (e.g., topography/elevation and sediment grain sizes), watershed hydrology (e.g., discharge data), and hydraulics (e.g., stage–discharge relationships, local flow velocity, and water depth measurements); (2) using third-party software (e.g., open TELEMAC-MASCARET, Hervouet & Ata, 2020) to select a mathematical model, define a spatiotemporal (implicit or explicit approach) discretization scheme (finite volume, finite differences, or finite elements), generating boundary conditions based on the input data, and implementing a numerical approximation and a method to solve the Navier–Stokes equations; (3) model calibration of relevant input parameters; (4) model validation to express model accuracy; and (5) a sensitivity analysis of model-specific and domain-specific parameters.

Model calibration of hydrodynamic parameters builds on measured hydraulic parameters such as flow velocity and water depth compared with modeled values. For example, if the modeled water depth is generally lower than the observed water depth, the roughness value to be used with the numerical model is iteratively adapted until a best global match between modeled and observed data is achieved (e.g., Barker et al., 2018; Carr et al., 2018). Morphodynamic

model parameters are, for instance, the Shields (1936) parameter (also referred to as critical dimensionless bed shear stress), which is vetted in the calibration process against observed elevation (i.e., topographic) change rates. For example, if the numerical model overestimates erosion, the Shields parameter is increased to emulate a higher resistance of grains against hydraulic forces. The outcome is a potentially over-calibrated hydro-morphodynamic model that needs to be validated with a second observation dataset in order to determine the model's accuracy (Beckers et al., 2020; Mosselman & Le, 2016). Both calibration and validation of hydro-morphodynamic models can be carried out using spatially explicit topographic data acquired at two different points in time. This time period is simulated in the numerical model by defining unsteady inflow conditions (i.e., a hydrograph) with sediment input at the upstream model boundary. Thus, the numerical model simulates erosion and deposition patterns that should possibly match observed topographic change. The observed topographic change for comparison with modeled topographic change stems from topographic data in the form of digital elevation models (DEMs). The precision and availability of DEMs have increased exponentially in recent years due to advances in remote sensing, such as airborne survey techniques, combined with ground-based and boat-based data recordings (Gruen, 2008; Haun & Dietrich, 2021; Marcus & Fonstad, 2008; Notebaert et al., 2009; Pajares, 2015). The comparison of two DEMs at two different times constitutes a DEM of difference (DoD) (Milan et al., 2007) with variable observation uncertainty that is accounted for in the procedure of producing a DoD (Wheaton et al., 2010). Eventually, the uncertainty in DoDs can significantly influence calculated sediment budgets over the observation period (Carley et al., 2012; Wheaton et al., 2010).

The comparison of observed topographic change in a DoD and numerically modeled topographic change is often done via pixel-by-pixel map comparisons or visual comparison (expert assessment). However, visual comparison is neither consistently repeatable (Shoarinezhad et al., 2020) nor transparent, and the pixel-by-pixel approach has shown to be overly sensitive to minor shifts (Power et al., 2001). Thus, state-of-the-art techniques compare numerically modeled and observed DoDs based on statistical correlation parameters resulting from a pixel-by-pixel (element-wise) comparison (Sutherland et al., 2004). A pixel-by-pixel comparison typically uses a gridded raster format of DoDs converted to numerical arrays and matches every modeled and observed array element (i.e., topographic change at an *ij*-array element). If both the modeled and the observed topographic change values are similar, pixel-by-pixel statistics indicate a high degree of correlation and therefore a high model accuracy and vice versa. Still, to the authors' best knowledge and in experience, hydro-morphodynamic models of natural river ecosystems are far from simulating with sufficient high precision to yield high correlation through pixel-by-pixel comparisons. For instance, sediment deposits or scour holes may be simulated slightly too far upstream or downstream than observed in reality. Such a model, simulating morphodynamic processes conceptually correctly, may yield as low a correlation with observed values in a pixel-by-pixel comparison as a completely random, non-representative model. This is why hydro-morphodynamic numerical models are often calibrated and validated based on expert assessments instead of statistical or stochastic methods to separate the wheat from the chaff. In this study, we introduce different levels of vagueness in the comparison of modeled and observed topographic change using fuzzy logic to attempt overcoming subjective expert assessment. To this end, and in the context of high uncertainty involved in hydro-morphodynamic modeling, we address the following research questions: (i) Are deterministic, hydro-morphodynamic numerical models performing better than a quasi-random, non-representative model, and how can the quality of different models be expressed? (ii) Can fuzzy logic aid in overcoming weaknesses of pixel-by-pixel map comparisons, and what are relevant fuzzification rules for map comparisons of modeled and observed data at different spatial resolutions?

To answer these questions, we adopt fuzzy map comparison methods that tolerate imprecision of topographic change in space (location) and magnitude (quantity). These methods were fundamentally first proposed by Hagen (2003) for comparing categorical geodata. First, we introduce fuzziness in space and preserve continuous numeric values of raster datasets, which is referred to as *fuzzy numerical map comparison* (Hagen, 2006). Second, we implement fuzzy sets and kappa κ statistics to establish a method for comparing categorical (i.e., categorized numerical values) raster maps. Thus, we assume that fuzzy sets theory addresses uncertainty and spatial-quantitative inaccuracies in a more explicit way than traditional methods to yield a better and objective appreciation of subjectively perceived correlation between two maps (in line with Pappenberger et al., 2007). To begin, we present a physical lab model of a shallow reservoir and a 6.9 km gravel–cobble bed river reach at the German–Austrian border, which both serve as test datasets in this study. Afterwards, we explain the implementation of fuzzy map comparison methods with novel features for the purpose of validating hydro-morphodynamic numerical models. The results and discussion outline the validity of the hypotheses and the sensitivity of particular parameters, and identify relevant fuzzy map comparison methods with reference to a quasi-random topographic change map.

2 | DATASETS

Two datasets serve for answering the above-defined research questions. The datasets involve a physical research model of a shallow reservoir (without specific scale) and a 6.9 km long gravel–cobble bed reach of the Salzach River between Germany and Austria. Both the shallow reservoir and the Salzach River were numerically modeled in previous studies and observed topographic change datasets are available.

2.1 | Physical model of shallow reservoir

Kantoush (2008) conducted a series of physical model experiments to analyze flow fields and sedimentation processes in shallow reservoirs with different shapes. We use one of the datasets from these experiments (referred to as T14 in the original publication) with a diamond-shaped reservoir. The experimental dataset includes maps of loose sediment deposition height from before and after a 7.5 h experimental run with steady inflow of 0.007 m^3/s and supply of suspended sediment in the form of crushed walnut shells (particle density of 0.0015 g/m^3) with an average diameter of 50×10^{-6} m and a sediment concentration of 3×10^3 g/m^3. A mixing tank upstream of the

experimental setup served for controlled supply of the water–sediment mixture. The supply rate was measured with a flow meter (precision ± 0.0001 m^3/s) and the sediment height was recorded at the end of the experiment with a mini echo sounder (precision $\pm 1 \times 10^{-3}$ m). At the beginning of the experiment considered, no sediment was present in the 4.0 m wide and 6.0 m long, diamond-shaped reservoir, which consisted of hydraulically smooth PVC walls and a flat (zero-slope) bottom. Figure 1a shows the topographic change recorded after the experimental run.

Shoarinezhad et al. (2021) reproduced the experimental run with a three-dimensional numerical model of the diamond-shaped physical model using the software SSIIM 2 (Sediment Simulation In Intakes with Multiblock, by Olsen, 2014) coupled with a calibration tool called PEST (Parameter Estimation and Uncertainty, by Doherty, 2015).

Both experimentally observed and numerically modeled topographic change datasets are used in this study with equidistances of 0.1 m (in x and y directions) between the points (resulting also in a pixel size of 0.1 m).

2.2 | The Salzach River

The Salzach River at the German–Austrian border is a heavily impaired mountain river that has been the subject of multiple studies in the framework of river restoration efforts (Beckers et al., 2018). This study uses datasets from previous works that include topographic surveys and numerical modeling of a 6.9 km long river reach (between 59.3 and 52.4 km upstream of the river's confluence with the Inn River). A section of the studied river reach is shown in Figure 2. The considered river reach experienced considerable channel erosion after flood events, with peak discharges of 3000 and 3530 m^3/s in 2002 and 2013, respectively. The annual average discharge is 240 m^3/s and a 100-year flood corresponds to a discharge of 3100 m^3/s. The considered river reach is characterized by an average slope of 0.1%, riverbed grain sizes range between 0.0005 and 0.14 m (corresponding to a global roughness of Manning's $n \approx 0.035$), and an average active channel width of approximately 82 m. The sediment transport regime is dominated by bedload (Beckers et al., 2020). Beckers et al. (2018, 2020) established a 2D hydro-morphodynamic numerical model to reproduce the two flood events and guide future sediment management efforts. In this study, we use numerical model outputs in the form of topographic change maps that stem from a 3-year simulation period (2010–2013), embracing only the large flood event in 2013. Topographic change maps were produced with the commercial *HYDRO_FT-2D* software (Nujić et al., 2019), which calculates topographic (active channel elevation) change along a quadrilateral grid with, on average, 16 m wide and 36 m long cells. Beckers et al. (2018) used an *expert*-calibrated model (i.e., input parameters were changed based on an expert assessment), and Beckers et al. (2020) presented the same model, but with a stochastic (Bayesian) calibration of input parameters. In this study, we test the fuzzy map comparison methods with results from both the expert assessment and the stochastically calibrated models.

2.3 | Baseline (placebo) map datasets

To assess model quality in the form of similarity between observed and model data, we generated quasi-random baseline rasters representing placebos for the results of the shallow reservoir and Salzach numerical models. The randomness in the baseline rasters stems from the BitGenerator of the *NumPy* library (Harris et al., 2020), where we

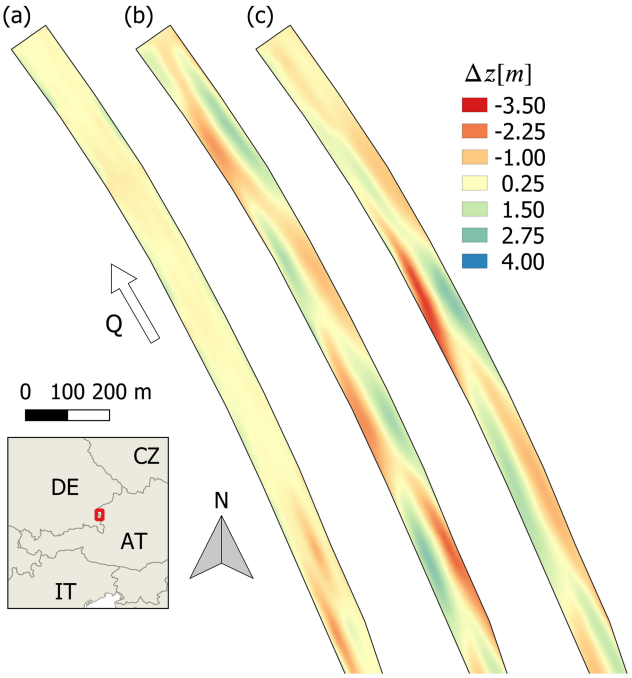

FIGURE 2 Downstream end of topographic change maps of the Salzach River between 2010 and 2013 (data source: Beckers et al., 2020): (a) observation map; (b) modeled map expert-calibrated; and (c) modeled map (stochastically calibrated) [Color figure can be viewed at wileyonlinelibrary.com]

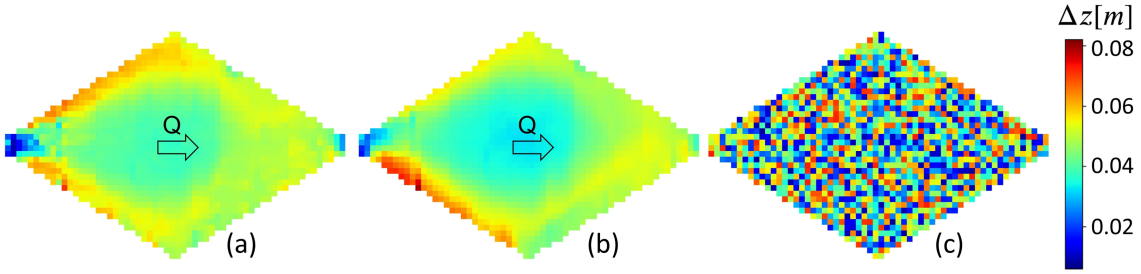

FIGURE 1 Topographic change maps of the diamond-shaped shallow reservoir with pixel size of 0.1 m (data source: Kantoush, 2008; Shoarinezhad et al., 2021): (a) observation map; (b) modeled map; and (c) quasi-random baseline map [Color figure can be viewed at wileyonlinelibrary.com]

defined a random value range according to the maps of observations (minimum and maximum observed topographic change in meters). Thus, provided that hydro-morphodynamic forces drive topographic change pattern, similarity measures between the baseline rasters and the observation rasters should be close to zero. Figure 1 visually compares the observed (Figure 1a) and modeled (Figure 1b) topographic change maps with the baseline placebo map (Figure 1c) of the shallow reservoir. In addition, Supporting Information Figure 9 in the supplemental material (C) visually compares the modeled topographic change maps with the baseline placebo map of the Salzach River. To our best appreciation and based on both figures comparing modeled and quasi-random, we assume that subjectively perceived correlation of the numerically modeled maps is higher than the correlation of the baseline maps and observation maps.

3 | METHODS

Instead of using a pixel-by-pixel comparison of modeled and observed data with statistical parameters such as the root mean square error (RMSE) or a correlation coefficient (e.g., Pearson's *r*), we introduce fuzzy logic to compare grid data in the form of maps. More information on model assessments with pixel-by-pixel (or value-by-value) statistics is provided in the supplemental material (A).

3.1 | Fuzzy logic

Depending on its value, a variable can be categorized as being true or false. For instance, if the variable *topographic change* is smaller than zero meters (i.e., negative), it can be attributed to belong to the category *erosion*. Thus, a curve to express degrees of truth (i.e., a *membership function*) of *erosion* can be plotted against all possible topographic change values and is a horizontal line at 1 (*True*) for topographic change values smaller than zero meters and 0 (*False*) for values larger than zero meters. Vice versa, we can define the category *deposition*. Thus, we attribute the belongingness (or membership) of topographic change values according to *Boolean logic* to a discrete *true–false* function. However, natural variables can cause more complex responses than just *true* or *false*, and with respect to topographic change we may want to consider an additional category of *little change*. Because a vertical line at a topographic change of zero meters is not representative (or quasi-impossible to happen), we define the *little change* category by a triangle-like function assigning real numbers between 0 (value does completely not belong to the category) and 1 (value does completely belong to the category). Similarly, we can modify the membership functions for the categories *erosion* and *deposition* so that the functions do not merely jump between 0 and 1 in the vicinity of zero meters topographic change, but rather fall linearly from 1 to 0 (*erosion*), or rise from 0 to 1 (*deposition*). Thus, a topographic change value close to zero meters can be assigned to multiple categories. The assignment of a variable's state to multiple categories through non-discrete (non-Boolean) functions corresponds to the definition of so-called *fuzzy sets* and represents the fundamental idea behind fuzzy logic (Zadeh, 1965; Zimmermann, 2011). Figure 3 exemplifies the attribution of a topographic change of 0.25 m to the categories of *deposition* and *little change*, which results in memberships of 0.5 in both categories. This attribution defines the *membership vector* μ of 0.25 m topographic change as $\mu = [\mu_{erosion} = 0.5, \mu_{little\ change} = 0.5, \mu_{deposition} = 0]$. We will compute membership vectors for every pixel in the fuzzy kappa method introduced below, which produces a map of *fuzzy pixels*. Thus, a *fuzzy pixel* represents multiple responses (in the form of fuzzy, non-Boolean pixel values) considering the response of a central pixel and its neighboring pixels. This is why the membership vector is also referred to as *fuzzy neighborhood vector* (Hagen, 2003).

3.2 | Fuzzy map comparison

A fuzzy map comparison of topographic change replaces pixel-by-pixel with fuzzy pixel comparisons to tolerate (numerical) modeling and/or measurement errors in space (location) and magnitude (quantity). According to the above-introduced definitions, the neighborhood of every fuzzy pixel is defined by a finite distance, referred to as the neighborhood radius *Rad* (in pixels), from a center pixel and incorporates all pixels within this distance (including the central pixel). Thus, the neighborhood constitutes a window of *N* pixels (Figure 4). The pixels within the neighborhood (neighbors and central pixel) are referred to as neighboring pixels.

To account for errors (or levels of data vagueness) in space (location) and magnitude, we consider two types of *fuzziness* in the neighborhood: (1) *fuzziness of location* and (2) *fuzziness of category*. Fuzziness of location means that a pixel in a modeled map may be similar to its exact equivalent pixel and to the neighbors of the equivalent pixel on a map of observation data. The degree to which a neighboring pixel contributes to the fuzzy representation of the considered (central) pixel is derived from a user-defined membership function. For a fuzzy map comparison, a distance decay membership function is preferable to a linear membership function (Figure 3) (Hagen, 2003). Figure 4 exemplifies a distance decay function defining the membership ω of a central pixel located at an *ij*-position on an array-like map, where *i* represents a row number (i.e., *y*-coordinate on a map) and *j* a column number (i.e., *x*-coordinate on a map). The figure also illustrates the pixel neighborhood defined by the neighborhood radius *Rad* that introduces spatial fuzziness. In the following, we refer to the position of the neighbors of central pixels by subscript ij, ι, where ι indicates positions relative to the central pixel. Thus, the distance decay membership function in Figure 4 involves the distance $d_{ij,\iota}$ between neighboring pixels (at ij, ι-positions) and the central pixel, and the so-called *halving distance* d_{halv}:

$$\omega(d_{ij,\iota}, d_{halv}) = \exp\left[\ln(1/2) \cdot d_{ij,\iota}/d_{halv}\right]. \quad (1)$$

The halving distance d_{halv} is the distance (in pixels) at which the membership function decays to half of its maximum value at the central pixel (Hagen, 2009).

Thus, fuzziness of location introduces a *spatial tolerance* in map comparisons, which is defined by $\omega(d_{ij,\iota}, d_{halv})$ and *Rad*, and constitutes fuzzy pixels. To avoid overestimating similarities between maps, a fuzzy pixel (e.g., in a modeled map) should not be compared to the analog fuzzy pixel of the map being compared (e.g., the map of observations) and vice versa. Consequently, a fuzzy pixel of one map is to be compared with the analog exact central (also called *crisp*) pixel of another (comparison) map (Hagen, 2003).

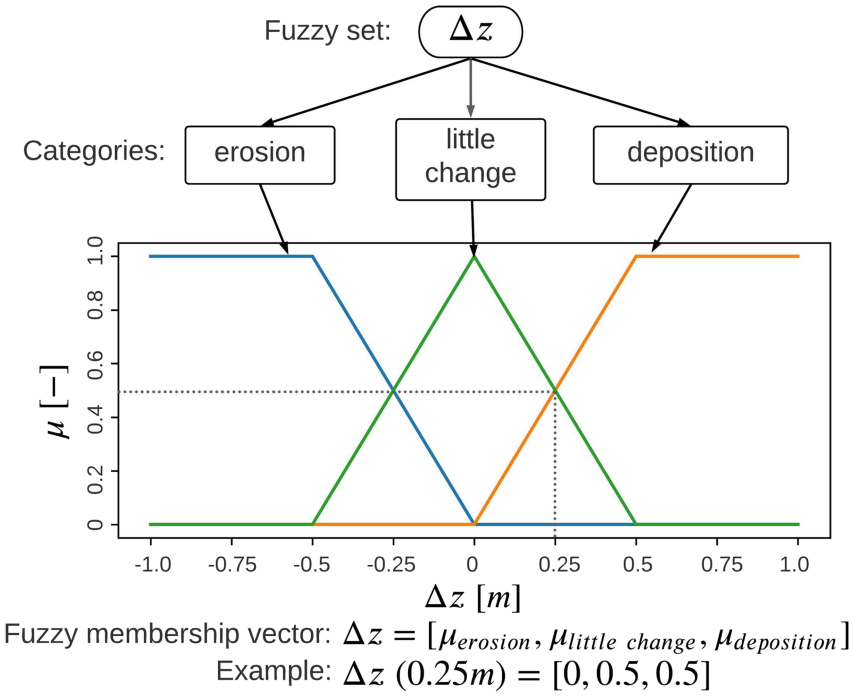

FIGURE 3 Definition of the three categories *erosion*, *little change*, and *deposition* through a fuzzy set and linear membership functions (in line with Zadeh, 1965) [Color figure can be viewed at wileyonlinelibrary.com]

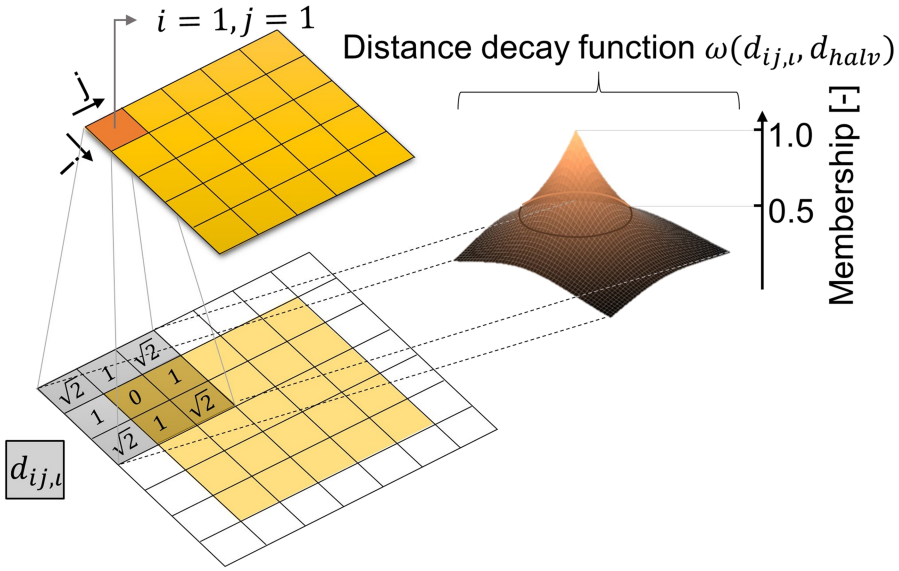

FIGURE 4 Qualitative example for introducing fuzziness of location through a fuzzy pixel with a neighborhood radius *Rad* of one pixel (total number of neighboring pixels $N = 9$) and an exponential distance decay membership function $\omega(d_{ij,\iota}, d_{halv})$ (Equation 1) with a halving distance d_{halv} of one pixel. $d_{ij,\iota}$ is the distance between the neighboring pixel and the central pixel at *ij* [Color figure can be viewed at wileyonlinelibrary.com]

In addition to fuzziness of location, fuzziness of category can be applied to categorical data such as *erosion*, *little change*, and *deposition*. Fuzziness of category recognizes that data categories may be similar to each other and expresses the similarity of categories in the form of a *similarity matrix* (see example in Supporting Information Table 8 in the supplemental material, G).

Both fuzziness of location and fuzziness of category can be integrated into map comparison techniques that involve different preprocessing of the original (input) map data. This study implements fuzziness of location in fuzzy numerical map comparisons and adds fuzziness of category in fuzzy kappa map comparisons.

3.2.1 | Fuzzy numerical map comparison

A fuzzy numerical map comparison applies to maps of real numbers only (i.e., non-categorical, continuous-valued raster maps of topographic change Hagen, 2006) and introduces spatial tolerance through fuzziness of location only. Before spatial tolerance is introduced, a fuzzy numerical map comparison starts with calculating the *per-pixel similarity* $\xi_{ij,\iota,om}$ between central pixel values of an observation map ($\nu_{o,ij}$) and the corresponding neighboring pixel values of a modeled map ($\nu_{m,ij,\iota}$) using Equation (2). Vice versa, the per-pixel similarity $\xi_{ij,\iota,mo}$ compares the central pixel values of a map with modeled values and neighboring pixels at *ij*, ι of a map with observed values (Equation 3):

$$\xi_{ij,\iota,om} = 1 - \frac{|\nu_{o,ij} - \nu_{m,ij,\iota}|}{\max(|\nu_{o,ij}|, |\nu_{m,ij,\iota}|)}, \tag{2}$$

$$\xi_{ij,\iota,mo} = 1 - \frac{|\nu_{m,ij} - \nu_{o,ij,\iota}|}{\max(|\nu_{m,ij}|, |\nu_{o,ij,\iota}|)}. \tag{3}$$

The per-pixel similarities $\xi_{ij,\,\iota,\,om}$ and $\xi_{ij,\,\iota,\,mo}$ are high (i.e., close to unity) when the observed and modeled pixel values are close to each other; in contrast, they are low (i.e., close to zero) when the observed and modeled pixel values differ significantly.

The next step introduces spatial tolerance by calculating a *per-pixel fuzzy similarity*$s^{fn}_{ij,om}$ between exact pixels in the observation map and fuzzy pixels in the modeled map (Equation 4). Vice versa, the per-pixel fuzzy similarity $s^{fn}_{ij,mo}$ between exact pixels in the modeled map and fuzzy pixels in the observation map is computed (Equation 5). Both $s^{fn}_{ij,om}$ and $s^{fn}_{ij,mo}$ represent the maximum of the product of the exact similarities $\xi_{ij,\,\iota,\,om}$ and $\xi_{ij,\,\iota,\,mo}$, respectively. Thus, the computation of $s^{fn}_{ij,om}$ and $s^{fn}_{ij,mo}$ involves an iterative calculation of the exact similarity of every central pixel of one map with the corresponding N pixels of the fuzzy pixel of the other map (ij, ι subscripts in Equations (4) and (5)):

$$s^{fn}_{ij,om} = \max^{N}_{\iota=1}\left[\xi_{ij,\iota,om}\cdot\omega(d_{ij,\iota},d_{halv})\right], \tag{4}$$

$$s^{fn}_{ij,mo} = \max^{N}_{\iota=1}\left[\xi_{ij,\iota,mo}\cdot\omega(d_{ij,\iota},d_{halv})\right]. \tag{5}$$

Next, the two-directional per-pixel similarity s^{fn}_{ij} is computed as the minimum of the two one-directional per-pixel fuzzy similarities $s^{fn}_{ij,om}$ and $s^{fn}_{ij,mo}$:

$$s^{fn}_{ij} = \min\left[s_{ij,om}, s_{ij,mo}\right]. \tag{6}$$

Ultimately, the global fuzzy numerical map similarity S^{fn} between an observation map and a modeled map is the average of the s^{fn}_{ij}-values of all (M) map pixels (Equation 7):

$$S^{fn} = \frac{1}{M}\sum_{ij} s^{fn}_{ij}, \tag{7}$$

where M denotes the total number of ij-pixels in a map; and superscript *fn* refers to the fuzzy numerical method (in contrast to the fuzzy kappa method *fk* presented in the next section).

3.2.2 | Fuzzy kappa map comparison

A fuzzy kappa map comparison combines fuzzy logic and Cohen's κ statistic to calculate the similarity between two categorical raster maps (Hagen, 2003, 2009; Hagen et al., 2005). Thus, the κ statistic accounts for relative observed agreement p_a between categorical datasets, which may occur by chance and partials out *expected agreement* p_e (Cohen, 1960):

$$\kappa = \frac{p_a - p_e}{1 - p_e}. \tag{8}$$

The subtraction of p_e from p_a (numerator) and its maximum possible value of 1.0 (denominator) yields a normalization of the agreement between two categorical maps, which is incorporated in the κ statistic (Cohen, 1960).

The expected agreement p_e depends on the size of the neighborhood and the relative frequency f of every category c in a map. The larger the neighborhood, the higher the probability that the central pixel matches one of the neighboring pixels. In the case of categorical map comparisons, p_e can be calculated as the sum of weighted matching probabilities of so-called *neighborhood rings* λ (Hagen, 2003):

$$p_e = \sum_{\lambda=0}^{R} p_\lambda \cdot \omega(d_\lambda, d_{halv}), \tag{9}$$

where R denotes the number of neighborhood rings, p_λ is the probability that the central pixels of both categorical maps match their counterparts (see calculation in the supplemental material, B), and d_λ is the distance between pixel centers in a neighborhood ring to the center of the central pixel. Thus, a neighborhood ring consists of pixels that have the same distance between their center and the center of the central pixel. For instance, the ring $\lambda = 0$ represents all pixels that have a distance of zero from a central pixel's center (i.e., the central pixel only). Similarly, the ring $\lambda = 1$ contains pixels with a distance of one pixel from the central pixel, but excludes neighboring pixels at the pixel corners, since their pixel centers have a distance of $\sqrt{2}$ from the center of the central pixel (see Figure 4).

The relative observed agreement p_a implies fuzziness of location and category, based on a fuzzy neighborhood vector containing the maximum values of a pixel's category membership $\mu_{ij,\,c}$. For instance, a pixel's fuzzy neighborhood vector $v_{ij,\,neigh}$ may contain values belonging to the membership in the above-introduced categories of *erosion* ($c = e$), *little change* ($c = lc$), and/or *deposition* ($c = d$):

$$v_{ij,neigh} = \left[\mu_{ij,e},\, \mu_{ij,lc},\, \mu_{ij,d}\right], \tag{10}$$

where the category memberships $\mu_{ij,\,c}$ are computed for every category c as the maximum of the product of the distance decay function $\omega(d_{ij,\,\iota}, d_{halv})$ and the categorical memberships $\mu_{ij,\,\iota,\,c}$ of neighboring pixels (with position index ij, ι) given by the similarity matrix:

$$\mu_{ij,c} = \max^{N}_{\iota=1}\left[\mu_{ij,\iota,c}\cdot\omega(d_{ij,\iota},d_{halv})\right]. \tag{11}$$

Thus, $v_{ij,\,neigh}$ represents the fuzzy union of all categorical memberships of the neighboring pixels ij, ι, weighted by the distance decay membership function. An illustrative example of the computation of $v_{ij,\,neigh}$ is provided in Supporting Information Figure 10 and Table 5 in the supplemental material (F).

Similar to the fuzzy numerical map comparison, the fuzzy kappa method also involves calculating a two-directional similarity per pixel. Therefore, for every pixel, its exact value in the observation map is used with the corresponding fuzzy pixel in the modeled map to calculate $s^{fk}_{ij,om}$, and vice versa to calculate $s^{fk}_{ij,mo}$. The superscript *fk* refers to the fuzzy kappa method (in contrast to *fn*, referring to fuzzy numerical similarity). The fuzzy pixel is represented by the fuzzy neighborhood vector, while the exact pixel is represented by a membership vector $v_{ij,\,bool}$, which is a vector of Boolean match or no-match values. For instance, $v_{ij,\,bool}$ for a pixel of the *little change* category can be calculated as follows:

$$v_{ij,bool} = \left[\mu^{bool}_{ij,c=e}, \mu^{bool}_{ij,c=lc}, \mu^{bool}_{ij,c=d}\right] = [0,1,0]. \tag{12}$$

Similar to the fuzzy numerical comparison technique, the two-directional per-pixel similarity s_{ij}^{fk} is calculated as the minimum of $s_{ij,om}^{fk}$ (similarity between the observation map and the modeled map) and $s_{ij,mo}^{fk}$ (similarity between the modeled map and the observation map):

$$s_{ij}^{fk} = \min\left(s_{ij,om}^{fk}, s_{ij,mo}^{fk}\right), \tag{13}$$

where $s_{ij,om}^{fk}$ and $s_{ij,mo}^{fk}$ are the maximum (i.e., the highest assigned) categories resulting from the minimum of the fuzzy pixel and the exact pixel values:

$$s_{ij,om}^{fk} = \max\left[\min_c\left(v_{ij,neigh,m,c}, v_{ij,bool,o,c}\right)\right], \tag{14}$$

$$s_{ij,mo}^{fk} = \max\left[\min_c\left(v_{ij,neigh,o,c}, v_{ij,bool,m,c}\right)\right]. \tag{15}$$

The global fuzzy numerical map similarity S^{fk} between an observation map and a modeled map is the average of the s_{ij}^{fk}-values of all (M) map pixels:

$$S^{fk} = \frac{1}{M}\sum_{ij} s_{ij}^{fk}. \tag{16}$$

To calculate the fuzzy kappa coefficient as a similarity measure between an observation map and a map of modeled topographic change, p_a is substituted in Equation (8) by S^{fk}:

$$\kappa_{fuzzy} = \frac{S^{fk} - p_e}{1 - p_e}. \tag{17}$$

In the case of perfect agreement between both observation and modeled maps, κ is 1.0, and it is 0.0 (or negative) if there is no agreement (Cohen, 1960). Note that the use and interpretation of the κ statistic are controversial because of original application limitations to dichotomous (binary) data types only (e.g., Kraemer, 1980; Maclure & Willett, 1987). A discussion on the κ statistic with illustrative explanations is provided with the supplemental material (E, Supporting Information Table 4).

3.3 | Hypotheses

To answer the research questions (i–ii) defined in the Introduction, we gradually test a set of hypotheses. The two fundamental hypotheses are:

I Any statistic identifies a numerical model being a better predictor for observed topographic change than the above-introduced quasi-random baseline model rasters.

II Fuzziness of location and/or fuzziness of category enable a more representative map comparison than other pixel-by-pixel methods.

The second hypothesis needs to be unpacked into two aspects: (a) the representation of subjectively perceived model performance; and (b) the capacity of compensating for spatial offset and/or magnitude errors of numerical models. Assuming that the perceived correlation corresponds to the actual correlation, these two aspects (a and b) conflict with each other because fuzzy similarities may identify correlation (i.e., $S^{fn} > 0$ and $\kappa_f > 0$) solely due to the introduced fuzziness that does not stem from the model performance. Thus, we investigate the truth of both aspects in this study by untangling hypothesis II into two sub-hypotheses that mutually exclude the truth of each other (i.e., verifying one sub-hypothesis implicitly falsifies the other sub-hypothesis):

IIa Fuzzy map comparisons reproduce subjectively perceived similarity (i.e., model performance) as a function of objective, quantifiable parameters. Thus, the ratio between statistics comparing numerically modeled raster maps with observation maps and statistics comparing baseline rasters with observation maps must be higher in the case of fuzzy similarities than in the case of pixel-by-pixel statistics.

IIb Fuzzy map comparisons compensate errors in location and/or magnitude as opposed to pixel-by-pixel statistics, which can lead to a statistic as low as for a non-sense model. Thus, the ratio between statistics comparing numerically modeled raster maps with observation maps and statistics comparing baseline rasters with observation maps must be lower in the case of fuzzy similarities than in the case of pixel-by-pixel statistics.

To test the two sub-hypotheses, we calculate the ratio ϱ as the fraction of similarities between numerically modeled and observation maps (numerator), and baseline raster and observation maps (denominator):

$$\varrho(\mathrm{STAT}) = \frac{\mathrm{STAT}_{model}}{\mathrm{STAT}_{base}}, \tag{18}$$

where STAT will be substituted by the tested pixel-by-pixel similarities (RMSE and Pearson's r) and fuzzy similarities (numerical and kappa). Note that since the RMSE is an error metric (the lower the value, the higher the model accuracy), the ϱ is to be interpreted inversely from the other ratios in the verification of the hypotheses. Thus, if ϱ ($1/\varrho$ for the RMSE) takes higher values for pixel-by-pixel similarities than for fuzzy similarities, we will reject hypothesis IIa and accept hypothesis IIb for a fuzzy map comparison method. For the converse, we will accept hypothesis IIa and reject hypothesis IIb.

3.4 | Algorithmic approaches

This study features the development of a novel algorithm for the validation of hydro-morphodynamic numerical models (available as an open-source Python package under the name Fuzzycorr; Negreiros,, (2020)). The algorithm pre-processes geospatial data to estimate the goodness of a modeled topographic change map compared with an observed topographic change map and comes with new routines for calculating fuzzy numerical map similarity (see supplemental material, D). In a first step, the algorithm transforms irregularly spaced topographic points to regular data grids in the form of raster files, which is necessary because of artifacts of measurements (e.g., because of different boat speed during sonar surveys) and the nature of numerical meshes (e.g., irregular triangular meshes). The

interpolated points are placed in the upper left corner of raster pixels using the SciPy interpolation library (Virtanen et al., 2020) and a user-defined interpolation method out of the following three options: linear, cubic polynomial, and nearest-neighbor interpolation. Here, we used a simple linear interpolation for the shallow reservoir testbed (Figure 1) and a cubic polynomial interpolation for the Salzach River (Supporting Information Figure 9 in the supplemental material, C).

The algorithm runs fuzzy numerical map comparisons with the regular raster maps based on the above-shown equations, which are implemented in modularized Python3 functions. Along with a pixel size value that drives the cell (pixel) size of the regular grid (raster), the algorithm accepts a float value defining the halving distance for the distance decay membership function (Equation 1) as function arguments.

In addition to the fuzzy numerical method that is implemented in the novel algorithm, we also tested fuzziness of category with the fuzzy kappa method using the Map Comparison Kit (Visser & De Nijs, 2006). For this purpose, we categorized modeled and observed topographic change raster maps based on natural breaks (Jenks, 1967) into degrees of sediment erosion and deposition. The rationale behind this choice is that natural breaks avoid excessive categorization of data (i.e., it creates a reasonable set of erosion and deposition categories), while it minimizes information loss. Moreover, the natural breaks method outperforms other methods, such as quantiles, standard deviation, or equal interval, when the frequency distribution of data has clear differences between value frequencies (Toshiro, 2002).

The new algorithm also generates comparison maps, which indicate the performance of a numerical model spatially explicit, along with global similarity measures resulting from fuzzy numerical map comparison (global fuzzy similarity S^{fn}, Equation 7).

For generating the results, the algorithm helped in (1) calculating similarities between baseline rasters of quasi-random values (corresponding to a placebo in clinical trials) and observation rasters (from the shallow reservoir and the Salzach River); (2) computing fuzzy map comparison similarities and pixel-by-pixel similarities between numerical model output and observation rasters; and (3) performing a sensitivity analysis with respect to the functional arguments of pixel size, neighborhood radius *Rad*, and halving distance d_{halv}. The algorithm output eventually enables verification of the above-defined hypotheses with the help of the similarity ratio ϱ (Equation 18).

4 | RESULTS

4.1 | Baseline (placebo) similarities

For testing the hypothesis that fuzzy map comparison better accounts for subjectively perceived correlation than pixel-by-pixel statistics, the first required results are similarity statistics comparing the quasi-random baseline rasters with observed topographic change. Table 1 lists the resulting baseline pixel-by-pixel statistics (RMSE and Pearson's *r*; see the supplemental material, A), the global fuzzy numerical baseline similarity S^{fn}_{base} (Equation 6), and the baseline fuzzy kappa coefficient κ_{base} (Equation 17). We used a halving distance of $d_{halv} = 2$ pixels and a neighborhood radius of $Rad = 4$ pixels for the shallow reservoir, and $d_{halv} = 4$ pixels and $Rad = 8$ pixels for the Salzach River.

TABLE 1 Baseline pixel-by-pixel statistics (RMSE and Pearson's *r*) and fuzzy similarities (S^{fn}_{base} and κ_{base})

Testbed	$RMSE_{base}$ (m)	r_{base} (–)	S^{fn}_{base} (–)	κ_{base} (–)
Shallow reservoir	0.021	−0.009	0.660	0.002
Salzach River	1.367	0.001	0.245	0.003

With a possible value range for S^{fn} between $-\infty$ (low correlation) and 1.0 (high correlation), the computed S^{fn}_{base} of 0.660 for the shallow reservoir indicates some correlation between observed and quasi-random topographic change. The value of S^{fn}_{base} for the Salzach River is lower with $S^{fn}_{base} = 0.245$. Yet, in both cases S^{fn} is considerably higher than the pixel-by-pixel metrics of RMSE and *r*, which indicates that the fuzzy numerical map comparison identifies a considerable arbitrary similarity. κ_{base} is close to zero in both cases, which is expected not least because κ_{base} contains the expected agreement p_e (see Equation 9).

4.2 | Fuzzy similarities of the numerical model of the shallow reservoir

The fuzzy numerical map comparison (Figure 5a) indicates high similarity (maximum s^{fn}_{ij} of 0.999) between observed and numerically modeled topographic change, primarily toward the right part (in flow direction) and also in the center of the reservoir. The smallest similarities (minimum s^{fn}_{ij} of 0.408) occur at the liquid boundaries of the numerical model (inflow and outflow regions). However, this particular numerical model is known to not well reproduce topographic change in the vicinity of the liquid boundaries (Shoarinezhad et al., 2021). The fuzzy kappa map comparison (Figure 5b) yields a wider range of similarity values, from a minimum of $s^{fk}_{ij} = 0.177$ to a maximum of $s^{fk}_{ij} = 1.0$. The fuzzy kappa similarities align with the fuzzy numerical similarities toward the right part (in flow direction) and the center of the reservoir, but indicate much lower similarity toward the left part (in flow direction) of the reservoir. Moreover, the fuzzy kappa similarity does not reflect the known inaccuracy of the numerical model in the region of its liquid boundaries. The categories used in the fuzzy kappa map comparison (Supporting Information Table 6 in the supplemental material, G) correspond to the range of observed topographic change between $\Delta z = 0.028$ m and $\Delta z = 0.056$ m (Kantoush, 2008). Both the fuzzy numerical and the fuzzy kappa map comparisons were calculated with a neighborhood radius *Rad* of four pixels and a halving distance d_{halv} of two pixels.

Table 2 shows the calculated global fuzzy map similarities S^{fn} (fuzzy numerical) and κ_f (fuzzy kappa). The fuzzy kappa map comparison generally yields lower similarities than the fuzzy numerical map comparison. Yet, κ_f of 0.517 still suggests that the numerical model performs significantly better than the quasi-random baseline model ($\kappa_{base} = 0.002$). The fuzzy numerical map comparison yields S^{fn} of 0.908, which also indicates that the numerical model performs better than the quasi-random baseline model ($S^{fn}_{base} = 0.660$). The pixel-by-pixel statistics reaffirm the better performance of the numerical model compared with the baseline model.

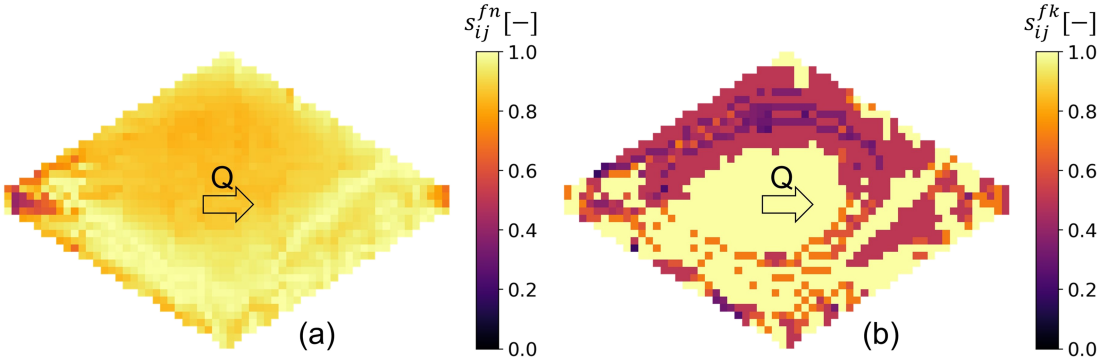

FIGURE 5 Comparison maps between modeled and observed topographic change: (a) fuzzy numerical and (b) fuzzy kappa map comparisons [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Pixel-by-pixel statistics (RMSE and Pearson's *r*) and fuzzy similarities S^{fn} and κ_f of the testbeds

Model	RMSE (m)	*r* (–)	S^{fn} (–)	κ_f (–)
Shallow reservoir	0.005	0.794	0.908	0.517
Salzach River (expert calibration)	0.806	0.109	0.347	0.054
Salzach River (stochastic calibration)	0.763	0.136	0.354	0.066

4.3 | Fuzzy similarities of the numerical model of the Salzach River

Figure 6 shows the result of the fuzzy numerical and fuzzy kappa map comparisons for the most downstream section of the study reach of the Salzach River between 2010 and 2013. Both map comparisons use a halving distance d_{halv} of four pixels (approximately 20 m) and a neighborhood radius *Rad* of eight pixels (approximately 40 m). The computed fuzzy numerical similarities s_{ij}^{fn} range between −0.070 and 0.997, indicating a significantly smaller similarity than in the case of the shallow reservoir. However, the small computed similarity corresponds to the subject perception of a bad per-pixel performance of the numerical model with respect to the topographic change maps in Figure 2. For instance, the stochastically calibrated model predicts a pronounced high erosion pattern in the left bank near the outflow boundary (Figure 2), while the observation map indicates sediment deposition. Negative fuzzy similarities are calculated in regions with a particularly low agreement between observed and modeled topographic change (Figure 6). With the computed per-pixel similarities s_{ij}^{fk} of the fuzzy kappa method ranging between 0.0 and 1.0, the combination of fuzziness of location and category rates the predictor capacity of the numerical models even lower, but still better compared to the quasi-random baseline model. s_{ij}^{fk} takes particularly low values in the vicinity of liquid model boundaries (generally below 0.2). The total range of observed topographic change at the Salzach River was between $\Delta z = -1.40$ m and $\Delta z = 1.68$ m, which defined the categories used in the fuzzy kappa map comparison (Supporting Information Table 7 in the supplemental material, G).

The global fuzzy numerical similarities S^{fn} of 0.347 for the expert-calibrated, and 0.354 for the stochastically calibrated Salzach River models subjectively suggest some similarity, for example, at the bottom of the map extents on Figure 2. The global fuzzy kappa similarities κ_f of 0.054 for the expert-calibrated and 0.066 for the stochastically calibrated Salzach River models indicate even less predictive capacities of the numerical models. However, the computed κ_f values

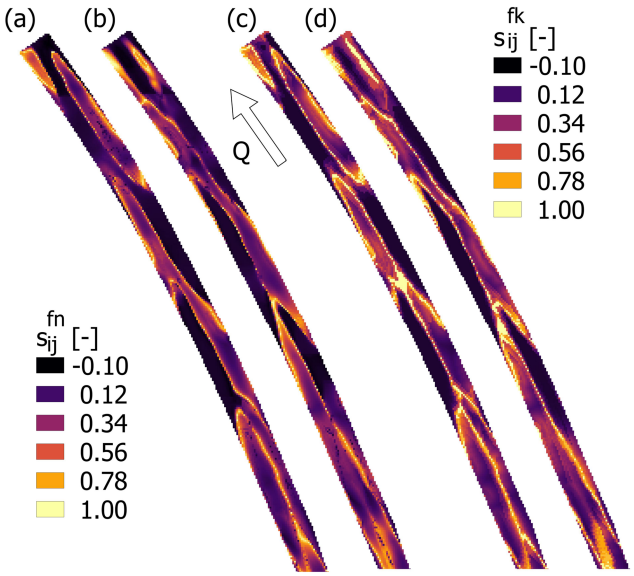

FIGURE 6 Map comparisons between modeled and observed topographic change of the Salzach River. Fuzzy numerical map comparison: (a) expert-calibrated model and (b) stochastically calibrated model; fuzzy kappa map comparison: (c) expert-calibrated model and (d) stochastically calibrated model [Color figure can be viewed at wileyonlinelibrary.com]

for the Salzach River models are still approximately 20 times higher than the expected fuzzy kappa agreement of $\kappa_{base} = 0.003$.

4.4 | Sensitivity analysis of geometric input parameters

The sensitivity analysis uses both testbeds (shallow reservoir and expert-calibrated Salzach River models) to evaluate S^{fn} (i.e., fuzziness of location only; Equation 7) as a function of pixel size, neighborhood radius *Rad*, and halving distance d_{halv}.

Pixel size is used in the very first step of the procedure for rasterizing irregular data and may affect the calculated similarities of fuzzy map comparisons. Figure 7 shows the pixel size variations with constant *Rad* of two pixels and a constant d_{halv} of one pixel. The resulting similarities do not show a significant variation for the shallow reservoir model with values ranging from 0.907 and 0.908. In contrast, S^{fn} increases by 36%, from 0.199 to 0.271, for the expert-calibrated Salzach River model. This difference between both observations is expected because the shallow reservoir inherently consists of grid data from laboratory experiments, in contrast to the naturally more complex Salzach River data with varying point density. The figures of the Salzach River model also show that S^{fn} increases with increasing pixel size. Thus, the increase in pixel size and, in consequence, the area where spatial vagueness applies causes an increase in the computed similarity of the Salzach River model.

For the sensitivity analysis of neighborhood radius *Rad* and halving distance d_{halv}, we varied *Rad* within the reasonable range of two (minimum to introduce fuzziness) and eight (maximum to avoid exaggerated fuzziness) pixels with a step size of two pixels (i.e., $N \in [2, 4, 6, 8]$). Similarly, we varied the halving distance between one and four pixels with a step size of one pixel (i.e., $d_{halv} \in [1, 2, 3, 4]$). Figure 8 plots the resulting global fuzzy similarity S^{fn} as a function of *Rad* and d_{halv}, where the pixel size is constant with 0.1 m for the shallow reservoir and 5.0 m for the Salzach River. In the case of the shallow reservoir, the effect of both *Rad* and d_{halv} is small (of the order of 10^{-4}), which can be attributed to a generally more homogeneous topographic change pattern (see Figure 1) than in the case of the Salzach River. The Salzach River was subjected to a more heterogeneous topographic change pattern in the modeled time period and increasing *Rad* and d_{halv} smooth the complexity of the topographic change pattern. Therefore, the similarity values S^{fn} increase with increasing *Rad* and d_{halv} by 73% for the Salzach River case, where the sensitivity of S^{fn} is higher with respect to increases in *Rad*. For instance, increasing *Rad* from two to four pixels (in the Salzach River case) yields an increase in S^{fn} of up to 46%, while an increase in *Rad* from 6 to 8 pixels yields a 4% increase only (where $d_{halv} = 4$). With very small neighborhoods of $Rad = 2$, variations in the halving distance d_{halv} are not significant. However, with increasing neighborhood radius, varying d_{halv} has significantly higher effects on S^{fn}. For example, when $Rad = 8$, increasing d_{halv} by one pixel (in the Salzach River case) yields an increase in S^{fn} of 12%. Thus, if fuzzy numerical mapping is to be optimized to compensate for spatial offset errors, the neighborhood radius *Rad* should be at least four pixels. While there is no such minimum value for the halving distance d_{halv}, its effect becomes more important in larger neighborhoods. In general, increasing *Rad* and d_{halv} will also increase S^{fn}.

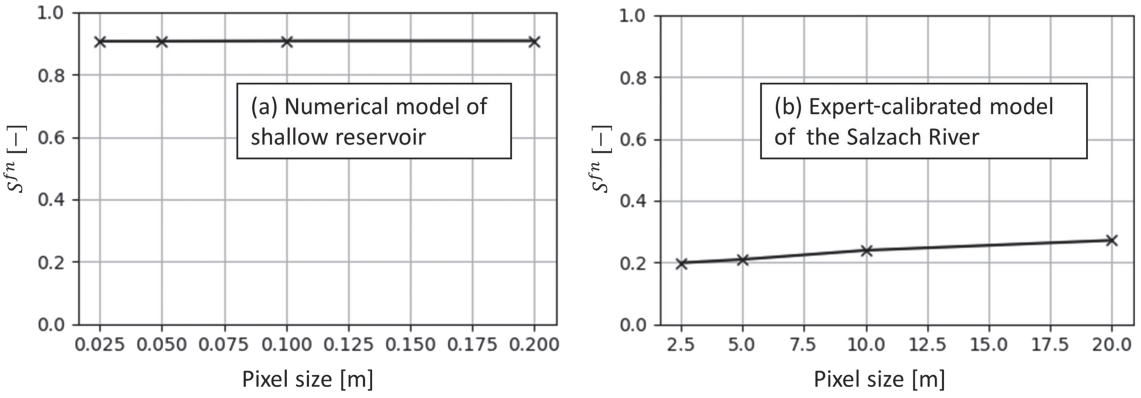

FIGURE 7 Effect of pixel size on the global fuzzy similarity S^{fn} of the testbeds

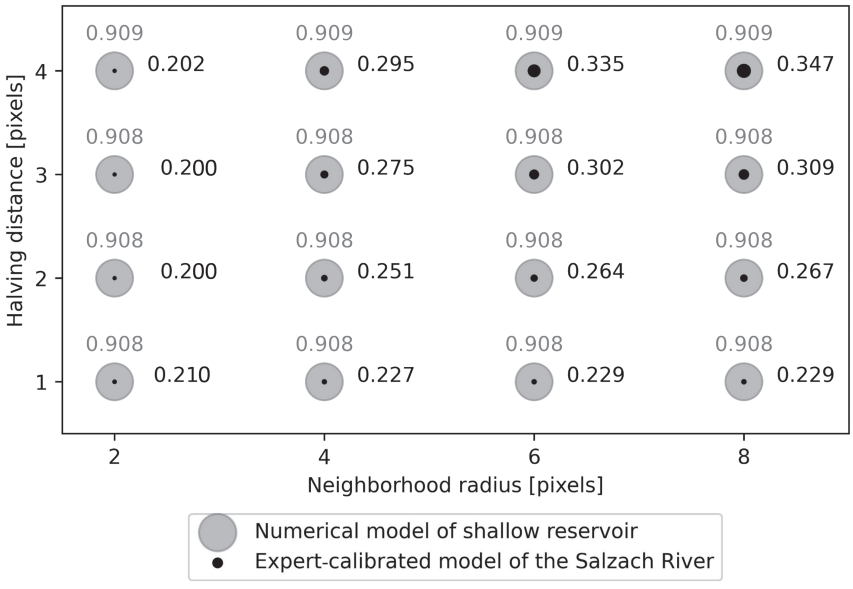

FIGURE 8 Global fuzzy similarity S^{fn} (Equation 7) as a function of neighborhood radius *Rad* and halving distance d_{halv}

TABLE 3 ϱ ratios (Equation 18) of calculated similarities for numerical model output and the baseline raster maps

Model	ϱ(RMSE)	$\varrho(r)$	$\varrho(S^{fn})$	$\varrho(\kappa_f)$
Shallow reservoir	0.24	88	1.38	259
Salzach River (expert calibration)	0.59	109	1.41	18
Salzach River (stochastic calibration)	0.56	136	1.44	22

4.5 | Validity of hypotheses

The validation of the hypotheses builds on the subjective observations that (1) the numerical model of the shallow reservoir reproduces well the measured values; and (2) the numerical models of the Salzach River reproduce alternating bank pattern well, but fail to predict exact topographic change rates. In consequence, hypothesis I requires that the similarities in Table 2 (numerical model compared with observations) generally suggest a better performance than the baseline similarities in Table 1. Thus, with respect to the ϱ ratio of similarities (Equation 18), ϱ(RMSE) must be less than 1.0 and all others should be significantly greater than 1.0 to verify hypothesis I. Table 3 lists ϱ for the similarity parameters used in this study and shows that both requirements (ϱ(RMSE) < 1.0 and all others >1.0) are fulfilled. Therefore, hypothesis I is true and we conclude that 2D numerical hydro-morphodynamic models have merit because they represent considerably better predictors than quasi-random placebo models.

The acceptance of either hypothesis IIa or IIb depends on the difference between ϱ of the pixel-b y-pixel and the fuzzy map comparisons. Table 1 shows that Pearson's *r* coefficient is 88 times higher for the comparison of the observation map and the modeled map than for the comparison of the observation map with the baseline raster of the shallow reservoir. The ratios are even higher for the Salzach River with ratios ϱ of more than 100. The RMSE of the comparison of the shallow reservoir's numerical model output divided by the RMSE of the comparison of the baseline raster should yield values close to zero to make RMSE a significant predictor. Compared with Pearson's *r*, the RMSE shows less amplification through the numerical model. In addition, the ϱ ratios of fuzzy numerical similarities are smaller than $\varrho(r)$ ratios of Pearson's *r*. The higher ϱ ratios for pixel-by-pixel comparisons do not support that fuzzy numerical comparisons yield a stronger representation of perceived correlation. However, $\varrho(\kappa_f)$ is much higher than $\varrho(S^{fn})$ and $\varrho(r)$, which indicates that fuzziness of category outweighs fuzziness of location in reproducing perceived map correlation. In addition, $\varrho(\kappa_f)$ is much lower for the more complex range of categories of the Salzach River (12 categories) than in the case of the shallow reservoir (six categories). These observations point out that fuzziness of category has a much stronger effect on similarity scores than fuzziness of location, with fewer categories yielding higher scores. Thus, while fuzziness globally leads to similarities closer to their maximum value of 1.0 than pixel-by-pixel comparisons (see Table 2), the verification of hypotheses IIa and IIb requires a distinction between fuzzy numerical and fuzzy kappa map comparisons.

With respect to fuzzy numerical map comparison, we observe that, according to hypothesis IIb, fuzzy numerical similarities tolerate model errors more than exact pixel-by-pixel comparisons (because of lower ϱ-values). In addition, the sensitivity analysis of pixel size, halving distance, and neighborhood radius showed that varying key input parameters plays a subordinate role (maximum amplification of 73%) compared to the map comparison method used. Consequently, hypothesis IIa is false and hypothesis IIb is true for fuzzy numerical map comparisons.

The hypothesis verification with respect to fuzzy kappa map comparison requires to account for the complexity of categories used, where, naturally, a smaller number of categories yields higher absolute (κ_f) and relative ($\varrho(\kappa_f)$) similarities than a higher number of categories. The implication for the validation of numerical models is that a fuzzy kappa map comparison with as few categories as possible (minimum of three categories for erosion, little change, and deposition) is best suited to indicate whether a numerical model produces fundamentally relevant results. However, when evaluating model performance to convey uncertainty, pixel-by-pixel statistics, especially Pearson's *r*, should be used as a global evaluation parameter. It follows that, for a small number of categories, hypothesis IIa is true and hypothesis IIb is false for fuzzy kappa map comparisons.

5 | DISCUSSION

5.1 | Perceived similarity

Verification of the hypotheses depends to some degree on the author's perception of observed similarity (e.g., in Figure 1). To the best of our ability, we objectively express perceived similarities here, and our observations are consistent in the case of the shallow reservoir. In the case of the Salzach River, however, the subjectively perceived similarity is not quite as straightforward because of the high heterogeneity in the observation data (see Figure 6 and Supporting Information Figure 9 in the supplemental material, C). However, to obtain a robust, quasi-objective opinion on the perceived similarity between datasets, ample surveys are required, which are in the domain of other disciplines beyond earth science. Here, we can only provide the outlook to verify our methods based on interdisciplinary approaches in the future.

5.2 | Data categorization

The categorization of real numbers of topographic change aims at assigning classes that meaningfully characterize erosion and deposition processes for fuzzy kappa map comparisons. Yet, there are known challenges in interpreting the kappa statistic because it ignores similarity between categories (Kraemer, 1980). For example, differences in categories such as *very little deposition* and *little deposition* are treated as being equally different like *much deposition* and *much erosion*. Moreover, the kappa statistic's main strength is in the comparison of dichotomous (binary *False* or *True*) data and it has known weaknesses in the comparison of nominal (i.e., multiple categories)

data (Maclure & Willett, 1987; Sun, 2011). As a result, the kappa statistic should not be used with manifold categories of either erosion or deposition. The similarity ratios ϱ (Table 3) highlight the decreasing relevance of the fuzzy kappa map comparison with an increasing number of categories (six categories for the shallow reservoir compared to 12 categories for the Salzach River). The rationale behind this categorization is to avoid subjective categorization by using the minimum and maximum of observation data and assigning intervals with natural breaks (Jenks, 1967). There are other, similarly objective possibilities for categorizing data, such as the average or the standard deviation of observation data to define numerical intervals for categories. However, because the interpretation of κ_f is challenging (Wealands et al., 2005) for multi-categorical data, it is only meaningful to use a small number of categories for topographic change map comparisons, such as the three categories of *erosion*, *little change*, and *deposition*, where the *little change* category should involve elevation differences corresponding to the accuracy of the DoD used. Only such a simplified categorization is useful for the assessment of a numerical model with the fuzzy kappa method, which consequently only enables us to confirm a rudimentary assessment of a numerical model. Yet, the fuzzy kappa map similarity is much more sensitive to a rudimentary working model compared to a placebo model (see Table 3) than in the case of pixel-by-pixel analysis. Therefore, the fuzzy kappa map comparison can represent an efficient feedback function for the validation of hydro-morphodynamic numerical models (see also discussion below).

5.3 | Alternative assessment methods

In addition to fuzzy map comparisons, skill metrics have been discussed by other authors to be more suitable to evaluate hydro-morphodynamic model performance than accuracy measures (e.g., RMSE) or correlation coefficients (e.g., Pearson's *r*) (Sutherland et al., 2004). A skill metric rates the performance of a model relative to the performance of a baseline (reference) prediction. To this end, a generic skill score suggests an improvement of model performance compared to a reference prediction and relative to the total possible improvement in performance (Murphy, 1988). Here, we additionally examine a novel skill score SS^{fn} that stems from the generic definition of a skill score (Murphy, 1988) and resembles the fuzzy kappa coefficient (Equation 17):

$$SS^{fn} = \frac{S^{fn} - S^{fn}_{base}}{1 - S^{fn}_{base}}. \quad (19)$$

Thus, the skill score SS^{fn} represents a performance statistic relative to the similarity of the baseline (placebo) rasters S^{fn}_{base} with similar interpretation ranges to Pearson's *r* (i.e., 0.0 means no correlation and 1.0 indicates perfect correlation). The skill score yields $SS^{fn} = 0.729$ for the shallow reservoir, $SS^{fn} = 0.135$ for the expert-calibrated model of the Salzach River, and $SS^{fn} = 0.144$ for the stochastically calibrated model of the Salzach River. Thus, the skill scores are slightly lower than Pearson's *r* for the shallow reservoir and slightly higher for the Salzach River models (see Table 2). Consequently, SS^{fn} seems to introduce a reasonable dampening of the correlation tendency under lab conditions and a reasonable amplification of correlation tendency under natural conditions.

5.4 | Relevance of map comparisons for numerical modeling

This study has its origin in the challenge of evaluating the performance of hydro-morphodynamic numerical models. With respect to the numerical model output compared with baseline (placebo) rasters (see Figure 1 and Supporting Information Figure 9 in the supplemental material, C) and the computed global fuzzy similarity values S^{fn} (Table 2), we verified the hypotheses that introducing fuzziness of location and category through fuzzy kappa map comparison aids to objectively mime subjective perception. The verification refers particularly to a low number of categories, which enables a reasonable expression of an expert's opinion (e.g., on the shallow reservoir model Shoarinezhad et al., 2021).

In the case of the Salzach River, the expert and stochastically calibrated numerical models suggest the development of a geomorphic landscape pattern corresponding to alternate bars. Because the considered reach of the Salzach River has undergone bank fixation, the pattern produced by the numerical model is reasonable and in good agreement with observations (Beckers et al., 2020; Chang, 1988). However, in reality sediment deposition and erosion did not exactly occur at the same place and to the same amount as predicted by the numerical models (see Figure 6 and Supporting Information Figure 11 in the supplemental material, H). Thus, the numerical models did correctly predict the geomorphic pattern, but they were imprecise with respect to exact topographic change rates. Yet, the pixel-by-pixel metrics of RMSE and Pearson's *r* state that the numerical models are close to completely wrong (*r* close to zero), but the fuzzy map comparisons slightly overcome the strict assessment of the pixel-by-pixel metrics through higher similarities (Table 2). In addition, the above-discussed skill score SS^{fn} (Equation 19) accounts for vagueness of location with the fuzzy numerical method and judges a numerical model in relation to a baseline reference model and can be considered a promising approach.

Finally, the geospatial assessment of model accuracy is a major advantage of fuzzy map comparison methods over global pixel-by-pixel statistics because fuzzy comparison maps indicate regions where the model performs particularly well or bad (e.g., visible in the comparison maps in Figures 5 and 6). Thus, fuzzy map comparisons are suitable for the validation of numerical models, but model calibration will need further development with feedback loops to optimize model performance. Such feedback loops will need to communicate with the input parameters of a numerical model and adapt these parameters based on particular computation output for every pixel. In this process, the model output can be converted to rudimentary, spatially explicit feedback responses with a fuzzy comparison map that reflects spatially explicit model improvements. For instance, a comparison map enables the calibration of spatially explicit variables such as roughness coefficients in an automated iterative process, which provides new opportunities compared to existing (half-)automated iterative calibration tools (e.g., Beckers et al., 2020; Doherty, 2015) for adjusting global model parameters, such as the Shields (1936) parameter.

Further research is needed to compare the novel performance statistics ($\varrho(\kappa_f)$ and $\varrho(S^{fn})$) with other purely data-driven approaches, such as machine learning methods, where decision making is based on numbers rather than intuition. In addition, different and more categories for landforms (e.g., crests, depressions, or plains) may lead to a more differentiated performance assessment of numerical models of reservoirs. In particular, a categorization of landforms based on standardized landform classification rules (e.g., with the LANDFORM algorithm from Klingseisen et al., 2008) has the potential to provide novel insights into numerical model performance assessment techniques, as an alternative to the natural breaks of sediment erosion and deposition categories used here.

6 | CONCLUSIONS

By means of map comparison methods, this study concludes that 2D hydro-morphodynamic numerical models are valuable predictors for the morphological evolution of river ecosystems. The viability of fuzzy logic in map comparisons is verified in the form of a fuzzy numerical map comparison method (fuzziness of location only) and a fuzzy kappa map comparison method (additional fuzziness of category). Introducing fuzziness of location generously compensates for model errors with respect to spatial offset of topographic change prediction over several years when the size of a fuzzy pixel is at least four exact (crisp) pixels. Fuzziness of category enables us to objectively reflect subjectively perceived correlation between observed and modeled topographic change, but only if a small number of categories (≤6) is used. Thus, a fuzzy kappa map comparison compensates less for model errors than a fuzzy numerical map comparison, but better expresses subjectively perceived correlation. In general, a strength of fuzzy map comparisons is the generation of comparison maps that spatially explicitly measure model quality (i.e., per pixel). In the future, the implementation of spatially explicit fuzzy-based map comparison can leverage the development of efficient methods for (semi-)automated hydro-morphodynamic numerical model calibration.

ACKNOWLEDGEMENTS

We thank Alex Hagen-Zanker, Vahid Shoarinezhad, and Felix Beckers, who provided us with helpful comments on the usage of fuzzy approaches for map comparisons and numerical model outputs. We also thank two anonymous reviewers for their constructive feedback, which helped us to improve this paper.

DATA AVAILABILITY STATEMENT

The Salzach River datasets were made available by Beckers et al. (2020). The numerically modeled topographic change dataset of the shallow reservoir stems from Shoarinezhad et al. (2021). Giovanni De Cesare (Operational Director of the EPFL PL–LCH) provided permission for using the laboratory datasets of experimentally observed topographic change from Kantoush (2008). The codes created for this study are available on Github (https://github.com/beatriznegreiros/fuzzycorr).

ORCID

Beatriz Negreiros https://orcid.org/0000-0001-7427-8871
Sebastian Schwindt https://orcid.org/0000-0002-7206-0542
Stefan Haun https://orcid.org/0000-0002-8202-4633
Silke Wieprecht https://orcid.org/0000-0001-6776-2446

REFERENCES

Barker, J.R., Pasternack, G.B., Bratovich, P.M., Massa, D.A., Wyrick, J.R. & Johnson, T.R. (2018) Kayak drifter surface velocity observation for 2D hydraulic model validation. *River Research and Applications*, 34(2), 124–134. https://doi.org/10.1002/rra.3238

Beckers, F., Noack, M. & Wieprecht, S. (2018) Uncertainty analysis of a 2d sediment transport model: An example of the lower river salzach. *Journal of Soils and Sediments*, 18(10), 3133–3144.

Beckers, F., Heredia, A., Noack, M., Nowak, W., Wieprecht, S. & Oladyshkin, S. (2020) Bayesian calibration and validation of a large-scale and time-demanding sediment transport model. *Water Resources Research*, 56(7), e2019WR026966.

Carley, J.K., Pasternack, G.B., Wyrick, J.R., Barker, J.R., Bratovich, P.M., Massa, D.A., Reedy, G.D. & Johnson, T.R. (2012) Significant decadal channel change 58–67 years post-dam accounting for uncertainty in topographic change detection between contour maps and point cloud models. *Geomorphology*, 179(Supplement C), 71–88. http://www.sciencedirect.com/science/article/pii/S0169555X12003819

Carr, K.J., Tu, T., Ercan, A. & Kavvas, M.L. (2018) Evaluating the applicability of a two-dimensional flow model of a highly heterogeneous domain to flow and environmental management. *Journal of the American Water Resources Association*, 54(1), 184–197. https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.12602

Chang, H.H. (1988) *Fluvial processes in river engineering*. New York: Wiley-Interscience.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Doherty, J. (2015) *Calibration and uncertainty analysis for complex environmental models*. Brisbane: Watermark Numerical Computing.

Ganju, N.K., Brush, M.J., Rashleigh, B., Aretxabaleta, A.L., Del Barrio, P., Grear, J.S., et al. (2016) Progress and challenges in coupled hydrodynamic–ecological estuarine modeling. *Estuaries and Coasts*, 39(2), 311–332.

Gruen, A. (2008) Scientific–technological developments in photogrammetry and remote sensing between 2004 and 2008. In *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book*, 7, CRC Press, Boca Raton, FL, pp. 39–44.

Hagen, A. (2003) Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235–249.

Hagen, A. (2006) Comparing continuous valued raster data: A cross disciplinary literature scan, Technical report. Maastricht: Research Institute for Knowledge Systems (RIKS).

Hagen, A. (2009) An improved fuzzy kappa statistic that accounts for spatial autocorrelation. *International Journal of Geographical Information Science*, 23(1), 61–73.

Hagen, A., Straatman, B. & Uljee, I. (2005) Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, 19(7), 769–785.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., *et al.* (2020) Array programming with NumPy. *Nature*, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Haun, S. & Dietrich, S. (2021) Advanced methods to investigate hydromorphological processes in open-water environments. *Earth Surface Processes and Landforms*.

Hervouet, J.M. & Ata, R. (2020) Telemac2d User Manual: EDF-R&D. http://www.opentelemac.org/ V8P1.

Jenks, G.F. (1967) The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186–190.

Kantoush, S.A. (2008) Experimental study on the influence of the geometry of shallow reservoirs on flow patterns and sedimentation by suspended sediments. (Doctoral dissertation), École Polytechnique Fédérale de Lausanne. Thesis No. 4048.

Klingseisen, B., Metternicht, G. & Paulus, G. (2008) Geomorphometric landscape analysis using a semi-automated GIS-approach.

Environmental Modelling and Software, 23(1), 109–121. https://www.sciencedirect.com/science/article/pii/S1364815207000953

Kraemer, H.C. (1980) Extension of the kappa coefficient. *Biometrics*, 36(2), 207–216.

Maclure, M. & Willett, W.C. (1987) Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology*, 126(2), 161–169.

Marcus, W.A. & Fonstad, M.A. (2008) Optical remote mapping of rivers at sub-meter resolutions and watershed extents. *Earth Surface Processes and Landforms*, 33(1), 4–24.

Milan, D.J., Heritage, G.L. & Hetherington, D. (2007) Application of a 3d laser scanner in the assessment of erosion and deposition volumes and channel change in a proglacial river. *Earth Surface Processes and Landforms*, 32(11), 1657–1674.

Mosselman, E. & Le, T.B. (2016) Five common mistakes in fluvial morphodynamic modeling. *Advances in Water Resources*, 93, 15–20. http://www.sciencedirect.com/science/article/pii/S0309170815001761

Murphy, A.H. (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116(12), 2417–2424.

Negreiros, B. (2020) *fuzzycorr*: GitHub. https://github.com/beatrizneg reiros/fuzzycorr (release version 0.0.1).

Notebaert, B., Verstraeten, G., Govers, G. & Poesen, J. (2009) Qualitative and quantitative applications of lidar imagery in fluvial geomorphology. *Earth Surface Processes and Landforms*, 34(2), 217–231.

Nujić, M., Hydrotech & Hunziker, Z.P. (2019) HYDRO_FT-2D reference guide (version 5.1.6): Add-ons to HYDRO_AS-2D for simulating sediment transport. Computer software manual. Aachen, Germany.

Olsen, N. (2014) A Three-Dimensional Numerical Model for Simulation of Sediment Movements in Water Intakes with Multiblock Option: User's Manual (Tech. Rep.). Trondheim, Norway: Norwegian University of Science and Technology (NTNU).

Pajares, G. (2015) Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVS). *Photogrammetric Engineering and Remote Sensing*, 81(4), 281–330.

Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R. & Matgen, P. (2007) Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrology and Earth System Sciences Discussions*, 11(2), 739–752. https://hal.archives-ouvertes.fr/hal-00305049

Power, C., Simms, A. & White, R. (2001) Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, 15(1), 77–100.

Shields, A. (1936) *Anwendung der Ähnlichkeitsmechanik und der Turbulenzforschung auf die Geschiebebewegung [Application of the similarity in mechanics and turbulence research on the mobility of bed load]*, Vol. 26. Berlin: Preußische Versuchsanstalt für Wasserbau und Schiffbau. http://resolver.tudelft.nl/uuid:61a19716-a994-4942-9906-f680eb9952d6

Shoarinezhad, V., Wieprecht, S. & Haun, S. (2020) Comparison of local and global optimization methods for calibration of a 3d morphodynamic model of a curved channel. *Water*, 12(5), 1333.

Shoarinezhad, V., Wieprecht, S., Kantoush, S. & Haun, S. (2021) The geometry effect of shallow reservoirs on the sedimentation pattern: Numerical analysis of lozenge-and hexagon-shaped reservoirs. In: *Proceeding of the 6th IAHR Europe Congress, Theme HS1, Reservoir Sedimentation: processes and management strategies*. Warsaw, Poland, pp. 661–612.

Sun, S. (2011) Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methodology*, 11(3–4), 145–163.

Sutherland, J., Peet, A.H. & Soulsby, R.L. (2004) Evaluating the performance of morphological models. *Coastal Engineering*, 51(8–9), 917–939.

Toshiro, O. (2002) *Classification methods for spatial data representation*. London: Center for Advanced Spatial Analysis, University College London. https://discovery.ucl.ac.uk/id/eprint/254/

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., *et al.* (2020) SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272.

Visser, H. & De Nijs, T. (2006) The map comparison kit. *Environmental Modelling and Software*, 21(3), 346–358.

Wealands, S.R., Grayson, R.B. & Walker, J.P. (2005) Quantitative comparison of spatial fields for hydrological model assessment: Some promising approaches. *Advances in Water Resources*, 28(1), 15–32.

Wheaton, J.M., Brasington, J., Darby, S.E. & Sear, D.A. (2010) Accounting for uncertainty in DEMs from repeat topographic surveys: Improved sediment budgets. *Earth Surface Processes and Landforms*, 35(2), 136–156.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, 8(3), 338–353.

Zimmermann, H.-J. (2011) *Fuzzy set theory and its applications*. Berlin: Springer.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Negreiros, B., Schwindt, S., Haun, S. & Wieprecht, S. (2022) Fuzzy map comparisons enable objective hydro-morphodynamic model validation. *Earth Surface Processes and Landforms*, 47(3), 793–806. Available from: https://doi.org/10.1002/esp.5285