A novel unsupervised method for assessing mesoscale river habitat structure and suitability from 2D hydraulic models in gravel-bed rivers

Received: 25 July 2021 | Revised: 23 June 2022 | Accepted: 8 July 2022

DOI: 10.1002/eco.2452

RESEARCH ARTICLE

WILEY

A novel unsupervised method for assessing mesoscale river habitat structure and suitability from 2D hydraulic models in gravel-bed rivers

David Farò[1] | Katharina Baumgartner[2] | Paolo Vezza[3] | Guido Zolezzi[1]

[1]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

[2]Unit of Hydraulic Engineering, Faculty of Engineering Science, University of Innsbruck, Innsbruck, Austria

[3]Department of Environmental, Land and Infrastructure Engineering, Politecnico di Torino, Torino, Italy

Correspondence
David Farò, University of Trento, Department of Civil, environmental and mechanical engineering, Via Mesiano 77, 38123 Trento, Italy.
Email: david.faro@unitn.it

Funding information
Italian Ministry of Education, University and Research (MIUR) under the Departments of Excellence; Italian Ministry of Education, University and Research (MIUR) under the Departments of Excellence, Grant/Award Number: L.232/2016; Autonomous Province of Bolzano/Bozen, under the research project 'FHARMOR Fish Habitat in Alpine rivers: integrating monitoring, modelling and remote sensing, Grant/Award Number: 17/34

Abstract

Application of mesoscale habitat models in gravel-bed rivers is increasingly common for a variety of purposes, from ecological flow design, impact assessment and conservation programmes. Integration with 2D hydraulic modelling offers the potential for broader applicability of mesoscale habitat models, extending applications to larger streams and nonwadable flow conditions, when on-the-ground and in-stream surveys are challenging or even prohibitive. In this work, a novel fully unsupervised procedure that allows the segmentation of the river channel area at a given flow condition at a scale that is consistent with the mesoscale is presented. Further, it defines an objective methodology to choose segmentation parameters and thus an optimal segmentation, based on intrinsic spatial properties of the resulting regions. Segmentation parameters are objectively selected by minimising a Global Score, which is based on three metrics representing intrasegment homogeneity, intersegment heterogeneity and an optimal range of segment numbers based on an empirically defined mesoscale. Applications of the model are tested on two reaches of the multithread Mareta and meandering Aurino Rivers in South Tyrol (NE, Italy). Model outcomes are compared with ground mesohabitat surveys, and habitat suitability is then assessed for three fish species (marble trout, grayling and European bullhead). A high level of agreement is found when comparing model- and survey-based habitat suitability estimates, with an overall value of $R^2 = 0.91$. The proposed approach shows potential for application of the mesohabitat concept for large gravel-bed rivers and nonwadable flow conditions. By allowing habitat estimates at flow ranges that could not be surveyed in-stream, the approach facilitates applicability of mesoscale habitat models to nonwadable conditions and large streams. The workflow is river-independent and fully unsupervised, as it does not require calibration or subjective choices of segmentation parameters.

Significance statement

Quantifying suitable habitat for riverine fauna is increasingly used for ecological flows assessment, with the use of mesoscale habitat modelling approaches becoming more common in the past few decades. Existing mesoscale habitat modelling approaches

often rely on field surveys, which, however, become prohibitive at nonwadable flow conditions and in large streams. Here we develop and test against field data a fully unsupervised approach able to extract mesohabitats and their hydraulic characteristics from the outputs of 2D hydraulic models. Compared with existing approaches, our approach allows an automated segmentation of the wetted reach into mesoscale units through the implementation of an unsupervised optimality step in which optimal segmentation parameters are defined, and a final mesohabitat mosaic is selected. The methodology makes it easier to expand the applicability of mesoscale habitat modelling to a broader range of river sizes and conditions.

KEYWORDS
ecohydraulics, hydromorphological unit, in-stream habitats, river morphology, unsupervised segmentation

1 | INTRODUCTION

Habitat modelling is commonly used in rivers to predict spatial distributions of aquatic species and biological responses to hydromorphological pressures. The spatial scale at which river habitat is modelled discriminates between two major classes of habitat models. Traditional habitat models, such as the widely used PHABSIM (Ahmadi-Nedushan et al., 2006; Bovee, 1982) characterise in-stream habitat suitability at the 'microscale', which refers to small ($\sim 1\,m^2$ or lower) river elements, with size of the order of flow depth or of the numerical cell of hydraulic models, having homogeneous hydraulic (water depth and velocity) and substrate conditions. Over the past few decades, the use of the 'mesoscale' in habitat modelling has become more common, with a number of modelling framework in use, such as the Mesohabitat Simulation Model (MesoHABSIM, Parasiewicz, 2007; Vezza et al., 2014), MesoCASiMiR (Eisner et al., 2005) and the Mesohabitat Evaluation Model (MEM, Hauer et al., 2009, 2011). Mesoscale habitat models use a representation of the river in the form of mesoscale patches (Wegscheider et al., 2020). The definition of the mesoscale used within this manuscript considers it a size between the river element (or microscale, 0.01–1 m), and the reach-scale (0.1–10 km), as typically defined within hierarchical river system classifications (Brierley & Fryirs, 2000; Frissell et al., 1986; Gurnell et al., 2016).

Many of the existing mesoscale approaches require mapping the mesohabitat mosaic through field-based visual reconnaissance (Borsányi et al., 2004; Eisner et al., 2005; Parasiewicz, 2007; Vezza et al., 2014), while water depth and velocity are surveyed in-stream. This works well for small streams and wadable conditions. Applications in large rivers or for high discharges are, however, more challenging. Furthermore, construction of reliable habitat-streamflow rating curves can be time-consuming, as habitat surveys need to homogeneously cover the range of the stream flow regime that is relevant for the purpose of the habitat modelling application. Such range normally includes low and middle flow conditions and lies well below bedload initiating flows in coarse-bedded streams. Such challenges can be overcome by the implementation of 2D hydraulic modelling (as, e.g., done in Hauer et al., 2011; Parasiewicz et al., 2012; Wyrick et al., 2014) that allow to simulate depth and flow patterns at different discharges. Delineation of HMUs from the hydraulic description of the reach is typically done using algorithm-based classification approaches (Hauer et al., 2009; Legleiter & Goodchild, 2005; Tamminga & Eaton, 2018; van Rooijen et al., 2021; Wallis et al., 2012; Woodget et al., 2016; Wyrick et al., 2014). But the extraction of the mesohabitat mosaic based on the outputs of 2D depth-averaged hydraulic models still remains challenging, as no universal and transferable solution between river type, discharge and morphology could be developed so far (Wegscheider et al., 2020), limiting therefore the use of 2D hydraulic models in mesoscale habitat modelling and therefore its potential for broader applications.

Furthermore, a wide range of classification systems have been developed to define mesohabitats, which makes it challenging to compare existing studies with one another, and to define thresholds that have a general applicability. These units are either characterised by their physical (hydraulic and geomorphic) characteristics, such as physical biotopes (Padmore, 1998), surface flow types (Davis & Barmuta, 1989) or geomorphic units (Bisson et al., 1982; Frissell et al., 1986) or were defined by identifying substrate patches based on the aquatic community they sustain, such as the functional habitats (Harper et al., 1992). Loose links between these two descriptions have been shown in, for example, Kemp et al. (1999) and Newson and Newson (2000). In this work, we use the definition of mesohabitats given by Belletti et al. (2017) for which mesohabitats correspond in size and location to hydraulic or geomorphic units (Belletti et al., 2017; Borsányi et al., 2004; Hauer et al., 2011; Parasiewicz, 2007; Vezza et al., 2014), as they result from the interplay between morphology and hydraulics and hence can be viewed as the building blocks of the river channel morphology (Belletti et al., 2017; Frissell et al., 1986).

Early attempts to define hydraulic thresholds to discriminate between different HMUs found the Froude number among the most suited variable (e.g. Jowett, 1993; Wadeson & Rowntree, 1998). Moir and Pasternack (2008) found that the combined use of depth and

velocity allowed to better discern between unit types, since considerable overlaps in hydraulic characteristics are usually found, and different combinations of velocity and depth can result in similar Froude number ranges (Clifford et al., 2006). This led, for example, to the development of an objective depth-velocity-based classification system (Wyrick et al., 2014), in which expert-based, river-specific thresholds were defined. Since the aim of the classification was to identify stage-independent geomorphic units, it was suggested to carefully choose a reference flow, for which the hydraulics best represented the underlying geomorphological forms.

To account for stage-dependent changes in HMU distributions, a flow-dependent classification system applied to 2D hydraulic modelling outputs was presented in Hauer et al. (2009, 2011). In this approach, scores are assigned to each simulated grid point based on classes of depth, velocity and shear stress. Finally, each point is assigned to a specific HMU type. Six different HMUs are classified. The score classes are defined through expert-based knowledge, requiring therefore a river-specific calibration. The approach presented by Hauer et al. has several advantages: First, it allows a direct classification of the flow into a specific HMU type; second, once river-dependent and expert-based thresholds have been set, the implementation of the method is very straightforward, since each hydraulic point can be assigned directly to a specific class. However, defining the hydraulic thresholds can be quite subjective (e.g. Clifford et al., 2006), and there might be a need to calibrate thresholds over a range of discharges (e.g. Hauer et al., 2011), requiring to perform river surveys over a longer time span. Finally, the resulting units and patches might be quite fragmented, particularly in complex and heterogeneous flow environments, resulting in patches that might be closer in size to the microscale.

To overcome the need to calibrate any parameter threshold and to account for the fuzzy nature of boundaries among HMUs, Legleiter and Goodchild (2005) used fuzzy-c-clustering to segment the river flow into six classes. The resulting map yielded patches of flow belonging to these classes, surrounded by uncertainty zones representing transitions between the patches. The same methodology was used in, for example, Tamminga and Eaton (2018) to study the effects of a large flood on the geomorphologic reshifting of the channel. Shortcomings of the method are its sensitivity to data quality, since the unsupervised definition of classes is highly dependent on the input data. Furthermore, the number of classes needs to be defined a priori. Although some validity indices (e.g. Arbelaitz et al., 2013) can be used, the final result still depends on subjective choices by the operator.

By adding a spatial contiguity constraint to an agglomerative hierarchical clustering algorithm, van Rooijen et al. (2021) showed that the delineation of habitat patches could be improved compared with standard applications of clustering algorithms. In particular, they found that the extent of patches could be better delineated, the transition between patches was smoother and that contiguous patches could be better distinguished, particularly avoiding an overfragmentation of units. While representing an improvement over previous clustering-based approaches, it still requires the operator to choose the final number of patches.

In recent years, new approaches to delineate the mosaic of HMUs from airborne remote sensing surveys have been developed, using, for example, RGB (Rivas Casado et al., 2015, 2017) or multispectral (Demarchi et al., 2016) orthophotos to map in-stream mesohabitats. While having a lot of potential in terms of speeding up river surveys and allowing to survey nonwadeable conditions, some of the drawbacks of field-based surveys remain. In particular, the need to repeat surveys at a number of discharges throughout the year and to acquire in-stream measurements of water depth and velocity within each remotely mapped mesohabitat.

The present study addresses the above mentioned issues by proposing a novel fully unsupervised procedure that allows the segmentation of the river channel area at a given flow condition. Further, it defines an objective methodology to choose segmentation parameters and thus an optimal segmentation, based on intrinsic spatial properties of the resulting regions. While the methodology does not attempt to map and label specific HMU types (such as riffles, glides or pools), it is aimed at finding an optimal segmentation of the wetted reach, mimicking a number of key features of a HMU mosaic from a spatial landscape perspective. The resulting units will be relatively homogeneous in terms of hydraulic features, while being clearly distinguishable from surrounding units and will have a scale which is consistent with the mesoscale, which was empirically defined based on the classification system proposed by Belletti et al. (2017).

The validity of the proposed methodology is evaluated by comparing habitat suitability estimates for target fish species resulting from the segmentation with estimates based on ground-survey mapping of HMUs from two case studies in gravel-bed, Alpine Rivers. The proposed unsupervised approach overcomes the need to calibrate segmentation parameters, allowing potentially to apply the same methodology to different river types, morphologies and discharges.

2 | MATERIALS AND METHODS

2.1 | Development of the methodology

The methodology focuses on a river subreach, considered representative of a morphologically homogeneous reach (sensu Rinaldi et al., 2013). It analyses the spatially distributed outputs (depth and velocity) of a 2D hydraulic model run on the target subreach, from which it extracts a spatial distribution of mesohabitats and of their suitability for a target species and life stage. It is based on a four-step workflow, using the k-means clustering technique (Lloyd, 1982) to initially segment flow points into clusters. The two-dimensional flow velocity and water depth field in the wet channel area is then polygonised into segments corresponding to the same clustering type, representing segment seeds, which are finally merged together through a region growing algorithm, thus ensuring that all segments grow to a size that is consistent with the mesoscale. In the final step, an optimal segmentation of the channel flow is chosen, by minimising a Global Score (GS). The GS is the average of three normalised metrics: the area-weighted variance, a measure of intrasegment homogeneity;

Moran's I, a measure of intersegment heterogeneity of the regions; and the mesonumber, which defines a range of optimal region numbers.

To avoid confusion with common morphologic terminology, the following key definitions are implemented within this work. A 'segment' represents a unit resulting from the segmentation process. The term 'region' is used in accordance with the region growing algorithm. A 'hydromorphologically based unit' (or 'mesohabitat') is an area defined by relatively homogeneous hydraulic conditions which are defined by local morphological characteristics of the channel (e.g. shape, bed slope and macroroughness elements), having a size-consistent with the mesoscale and, although not equal to, have similar characteristics to the geomorphic units defined in Belletti et al. (2017). Each hydromorphologically based unit is composed by a set of smaller 'hydraulic units' (sensu Belletti et al., 2017), which are patches with homogeneous hydraulic conditions, smaller in size than mesohabitats, resulting from the first segmentation step (following steps 1 and 2 of the workflow) or during earlier region growing iterations (step 3).

The four segmentation steps (Figure 1) to be performed on the outputs of the hydraulic model are as follows: (1) k-means clustering of computational cells where the hydraulic model computes depth and velocity values, (2) polygonisation of hydraulic units resulting from step 1, (3) region growing and (4) selection of segmentation parameters, which define an optimal mosaic of hydromorphologically defined segments.

A detailed description of each individual step is presented below.

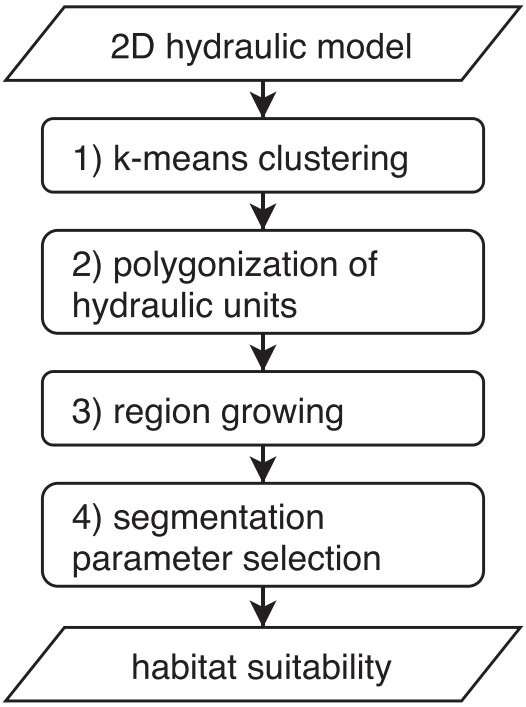

FIGURE 1 Schematic representation of the mesohabitat modelling workflow

2.1.1 | Step 1: K-means clustering

In this first step, the k-means clustering algorithm (Lloyd, 1982) is implemented to clusterise flow data. Clustering algorithms are considered to be unsupervised grouping algorithms, which work only based on a notion of similarity between the data. The k-means algorithm works by minimising the variance, that is, the within-cluster sum of squares, and maximising squared deviations between objects of different clusters, that is, the between-cluster sum of squares. The algorithm requires the choice of the number of clusters k, which can be either set a priori through, for example, knowledge of the real number of groups in the data or empirically validated a posteriori by comparing different choices of k with internal or external validation indexes (e.g. Arbelaitz et al., 2013). To improve performance and avoid local optimal clusterisation, the algorithm is initialised using the k-means++ approach (Arthur & Vassilvitskii, 2007). The flow data, consisting of depth and depth-averaged velocity values from the 2D hydraulic simulations, were normalised before clusterisation to have zero mean and unit standard deviation. For the selected k value, the algorithm is applied to the data set, and each cell of the computational domain of the hydraulic model, with associated depth and velocity value, is assigned to a cluster.

In this application focused on river habitat modelling, choosing a hydromorphologically consistent number of clusters should account for the reported diversity of HMUs, which vary depending on the channel morphology (Belletti et al., 2017), for which the order of magnitude is ~ 10. To account for this diversity, initialisation of the algorithm based on a broad range of k values is therefore recommended.

2.1.2 | Step 2: Polygonisation of hydraulic units

The mesh used in the hydraulic model is first rasterised into a regular grid of square cells of a meaningful size, chosen based on channel size and its morphological complexity. Based on the results from step 1, a cluster type is assigned to each raster tile. All contiguous tiles belonging to the same cluster type are then merged together. Once the merging is completed, the arising regions are polygonised. This second step serves the creation of a first segmentation, which is solely based on the clusterisation of step 1. Each segment is then used as seed for the region growing algorithm implemented in the following step 3.

2.1.3 | Step 3: Region growing

To ensure that all units have a size that is consistent with the size of mesohabitats in rivers (Belletti et al., 2017), a region merging algorithm is applied, which iteratively merges contiguous units with each other. This step is implemented to avoid the creation of too many small units, which are not consistent with the mesoscale. Region growing follows three main principles, which try to mimic the visual recognition and classification of HMUs in the field: (1) small merges

first, (2) smallest merges with largest and (3) similar merges with similar. At every iteration of the algorithm, a merging score is assigned to any potential merge between each region with its rook-case neighbouring regions. A first ranking is defined based on the first principle, according to which the regions are given highest priority if they are smaller in size. If more than one region has the same size, all regions of same size are ranked again according to principles 2 and 3.

Principles 2 and 3 are computed in the form of a similarity matrix, normalised in the range 0–1. The negative squared Euclidean distance $d(x_i,x_j) = \|x_i - x_j\|_2^2$ is used to compute similarity between regions, with x_i and x_j representing generic n-dimensional variables ($\vec{x_i} = (x_{i1},...,x_{1n})$ and $\vec{x_j} = (x_{j1},...,x_{jn})$), and $\left\|\vec{h}\right\|_2 = \sqrt{h_1^2 + ... + h_{n^2}}$ defining the Euclidean norm of a generic n-dimensional vector $\vec{h} = (h_1,...,h_n)$. For the second criteria, this distance is computed on the area of contiguous regions $d(A_i,A_j)$. To limit its range between 0 and 1, an exponential variant of the distance is used: $d_2 = exp(1/d(A_i,A_j))$. Similarity for criterion 3 is computed on the average depth and velocity values ($h_i = (d_i, v_i)$) of each region: $d(h_i,h_j)$. To prioritise regions which are similar and limit the range between 0 and 1, the following variant of the distance is used: $d_3 = 1 - \frac{d(h_i,h_j)}{min(d(h_i,h_j))}$. A combined criterion is then computed as the product of the two distances $d = d_2 \cdot d_3$.

After each iteration, the lowest ranked region will be merged to one of its neighbours, diminishing the overall number of segments from n to $n-1$. This procedure is then repeated, until the whole reach is merged into one final segment. All intermediate segmentation steps are saved, and will be evaluated for optimality in the following step (4).

2.1.4 | Step 4: Segmentation parameter selection

Resulting from the first three steps, a matrix of segmentation combination arising from the initial clusterisations (for a range of k values) and the final segment numbers (for a range of n values) is composed. A total of $k \cdot n$ combinations are computed. A GS is used to rate each segmentation combination and to select an optimal result. The GS is computed as the sum of three scores: the weighted averaged variance (v), a measure of intrasegment homogeneity; Moran's I (MI), a measure of spatial autocorrelation, quantifying intersegment heterogeneity; and the mesonumber (mn), an empirically based range of optimal segment number.

The weighted averaged variance (Espindola et al., 2006) is calculated as follows:

$$v = \frac{\sum_{i=1}^{n} A_i \cdot v_i}{\sum_{i=1}^{n} A_i} \quad (1)$$

where n is the total number of segments, A_i is the area and v_i is the variance of a variable (e.g. water depth or velocity) of each i-th segment. See Figure 2 for an illustration.

Moran's I (Fotheringham et al., 2000) is computed as follows:

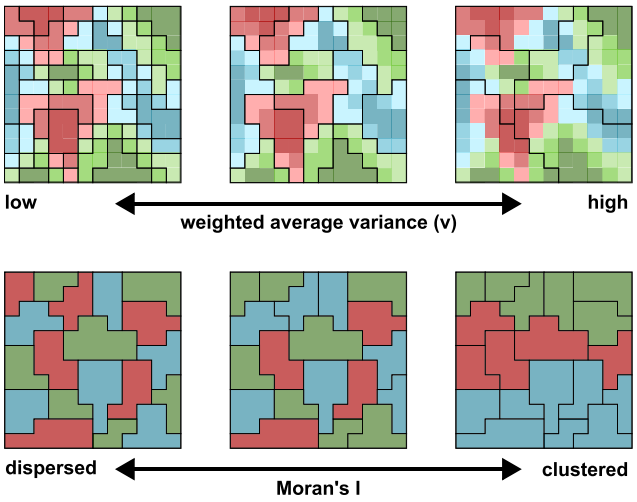

FIGURE 2 An illustration of weighted average variance and Moran's I

$$MI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i \neq j}\sum w_{ij}\right)} \quad (2)$$

with w_{ij} the contiguity matrix of regions i and j, which is set to 1 for regions that share a common boundary and 0 for nontouching regions; y_i represents the average value of a descriptive variable (e.g. water depth) of each region and $\bar{y}$ the average for the analysed reach of the same variable. Moran's I captures the degree of autocorrelation between adjacent units and will define how much on average each region is different from its neighbouring regions. Its value can range between -1 (a perfectly dispersed pattern) and 1 (high spatial correlation), with values close to 0 representing random patterns. See Figure 2 for an illustration.

Finally, the mesonumber follows a log-normal distribution:

$$mn = lognorm(\mu_{mn}, \sigma_{mn}) \quad (3)$$

with the mean μ_{mn} and σ_{mn}, respectively, the logarithmic mean and standard deviation of the optimal segment number for a given reach, respectively. These values are empirically based (see Section 2.1.5 for more details).

The individual measures of v, MI and mn are normalised to a common range from 0 to 1 according to the formula from Espindola et al. (2006):

$$x_{norm} = \frac{x_{max} - x}{x_{max} - x_{min}} \quad (4)$$

in which x_{max} and x_{min} represent the maximum and minimum values and x the variable to normalise. This defines the direction of optimisation as a minimisation of the scores.

The GS is then computed as the average of the three normalised scores:

$$GS = (v_{norm} + I_{norm} + mn_{norm})/3 \quad (5)$$

An optimal segmentation will be defined by small values of Moran's I (i.e. neighbouring regions are dissimilar), with low intra-segment variance (i.e. each region is internally homogeneous), and a number of regions consistent with an empirically derived mesoscale.

The GS here implemented used two indices (*v* and *MI*) that were introduced in Espindola et al. (2006) and are commonly used in the field of image segmentation (e.g. Kavzoglu et al., 2017; Räsänen et al., 2013; Ventura et al., 2018). A third score, the optimal mesonumber (*mn*), is instead ecohydraulically based, and was introduced to allow selection based on a third criterion, which favours selection of segmentations consistent with an empirically defined mesoscale of HMUs.

2.1.5 | Empirically derived parameters for segmentation

Both the area-weighted variance and Moran's I require to choose a variable (or a set of variables) to compute the variance. To define what variable is best suited for the implementation, the application of the GS was tested for the hydraulic variables depth, velocity, Froude number and the averaged variance of depth and velocity. Following the definition of optimal segmentation, the variable that most consistently minimised the GS for all case studies was chosen for the segmentation workflow. To allow comparability between GSs values from different rivers and discharges, the following normalisation for *v* and *MI* was used, as suggested in Böck et al. (2017). In the case of *v*, normalisation was computed according to the equation: $v_{norm} = v/var(x_{reach})$. Here $var(x_{reach})$ represents the variance for the variable x for the whole reach. In the case of Moran's I value, the minimum and maximum values were set to the range allowed by *MI*, −1 and 1.

Additionally, to define the distribution of the mesonumber (Equation 3), two parameters (i.e. the logarithmic mean μ_{mn} and standard deviation σ_{mn}) need to be computed. These values are based on empirical estimations of normalised sizes of surveyed HMUs, which we define here as the mesosize:

$$c_i = \frac{A_i}{\overline{W}^2} \quad (6)$$

The mesosize c_i represents the HMU area (A_i) normalised by the square of the average channel width $\overline{W}$, which is computed as the longitudinally weighted average of the width of all channels in the reach submerged by water. The log-normal distribution of *mn* (Equation 3) is derived from the log-normal distribution of the empirically based mesosizes c_i, corresponding to the distribution of potential mesohabitat numbers *N*, which can be estimated for any reach characterised by the wetted channel area A_{reach} and the channel width $\overline{W}$ following the function:

$$N_i \sim \frac{1}{c_i} \cdot \frac{A_{reach}}{\overline{W}^2} \quad (7)$$

Being c_i log-normally distributed, the resulting N_i will also be log-normally distributed. The parameters μ_{mn} and σ_{mn} that define the distribution of the mesonumber vary for every reach, since they depend on the values of the reach-specific area (A_{reach}) and of its average width ($\overline{W}$).

2.1.6 | From unsupervised mesohabitat extraction to suitability

Following step 4, an optimal segmentation (i.e. a mosaic of hydromorphologically defined units) for each modelled discharge Q is obtained. Application of a Habitat Suitability Model (HSM) to each mosaic allows estimating habitat suitability for a given target species and life stage at the mesoscale. By integrating estimates at different discharges, a habitat-flow rating curves can be derived (Vezza et al., 2014). It must be noted that only hydraulic habitat descriptions can be derived using a 2D hydraulic model, which must be integrated with further field or remote sensing-based mapping. Alternatively, simplified HSM which only account for hydraulic descriptors could be used.

2.2 | Application to case studies

2.2.1 | Study sites

To showcase the applicability of the presented methodology, the segmentation procedure was tested on two gravel-bed rivers located in the central north Italian Alpine area, in the region of South Tyrol, with different hydromorphological characteristics and channel size (Table 1). Both study reaches underwent recent river restoration projects, and in both of them, ground habitat surveys using the Meso-HABSIM methodology were performed, thus providing a reference against which to compare the outcomes of the proposed approach.

The River Mareta (also known under the German name Mareiterbach, Figure 3) has been heavily modified over the last century, by

TABLE 1 Hydromorphological characteristics of the study reaches on the Mareta and Aurino Rivers

Feature	Mareta River	Aurino River
Catchment area	206.53 km^2	607.86 km^2
Elevation	939.59 m a.s.l. (at hydrometric station)	815.77 m a.s.l. (at hydrometric station)
Slope	~1%	~0.4%
Reach morphology	Multithread	Meandering
Average wet channel width (reach)	~15 m	~26 m

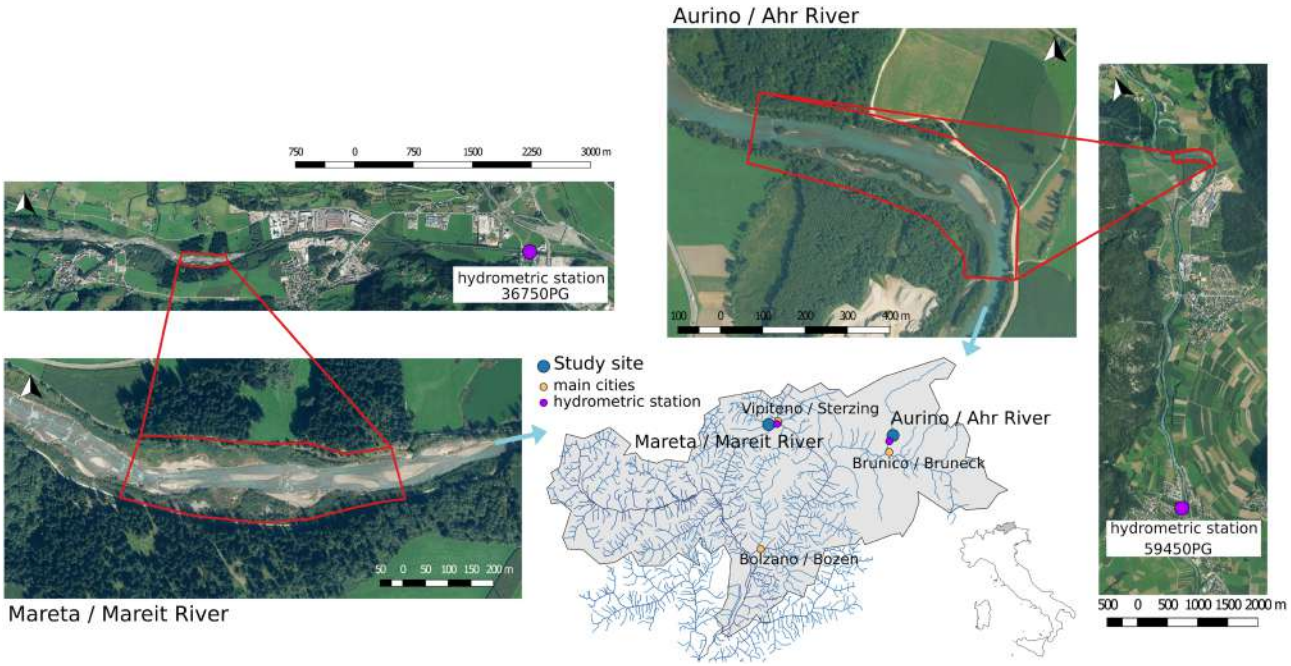

FIGURE 3 Maps of the case studies, two reaches of the Mareta and of the Ahr Rivers. The location of the region (Autonomous Province of Bolzano / Bozen) within Italy and of the study reaches (blue dots) within the province is shown. The study reaches are highlighted in red. The map shows also the hydrometric stations downstream of the study reaches (36750PG for the Mareta and 59450PG for the Aurino) that provided the hydrological time series

channelisation, construction of grade-control structures and intense phases of gravel mining, which caused changes in the active channel width and in its morphological pattern. Following works done in 2009, in which these grade-control structures and bank protections have been removed, a 3 km long reach located few kilometres upstream the city of Vipiteno/Sterzing was restored to a braided morphology (Zerbe et al., 2019). A 500 m subreach was selected as study reach.

The Aurino (also called Ahr River, Figure 3) mostly presents a single thread, sinuous to meandering morphology, which is laterally fixed by bank protection and levees. It has been partly restored starting from 2003 (Campana et al., 2014; Zerbe et al., 2019), with actions aimed mainly at improving riparian woodlands and fish habitats. A 750 m subreach upstream of the Gais municipality was selected as study site.

A comparison of the selected case studies in terms of their hydromorphological characteristics can be seen in Table 1 and Figure 4.

2.2.2 | Field data collection

During the time span ranging from 2016 to 2018, the mosaic of HMUs at varying discharges was repeatedly mapped at the study reaches according to the MesoHABSIM methodology (Parasiewicz, 2007; Vezza et al., 2014). A list of the main HMU types with a short description can be found in Table 2.

HMUs were mapped on the Mareta River using a portable GIS system connected to a TruPulse 360B laser rangefinder. The survey was georeferenced using the GPS system of the Getac PS336, running ESRI Arcpad 10.2.5. The HMU mosaic was recorded at three discharges, from low to middle flows, at $Q = 1.7$, 3.4 and 10.4 m^3/s.

Due to the larger size of the river reach, the Leica GS16, a handheld GPS-RTK was used on the Aurino River, by moving along the shore and delimiting the perimeters of each HMU. HMUs were mapped at two discharges, at low to middle flows of $Q = 7.5$ and 10.45 m^3/s.

The survey discharges and their percentages of exceeded time flow are highlighted on the Mareta and Aurino hydrographs and flow duration curves in Figure 4.

2.2.3 | 2D hydraulic modelling

Detailed and high-resolution bathymetries of the river reach have been acquired in December 2016 using a Airborne LiDAR Bathymetry scanning technology by the company AirborneHydroMapping GmbH (Innsbruck, AT). This work was carried out within the FHARMOR project (Farò et al., 2018). The acquired points have been georeferenced, classified and postprocessed to correct for refraction using Snell's law, and finally, a uniform mesh with grid widths of 50 cm was obtained (Baumgartner, 2020).

Hydraulic modelling of the studied reach was performed with the hydrodynamic model HYDRO_AS-2D (Nujić & Hydrotec, 2017), which numerically integrates the shallow water equations using a finite volume method. An optimal spatial distribution of roughness coefficients (Strickler) was then found by calibrating the model minimising the RMSE values between simulated and surveyed water levels at various

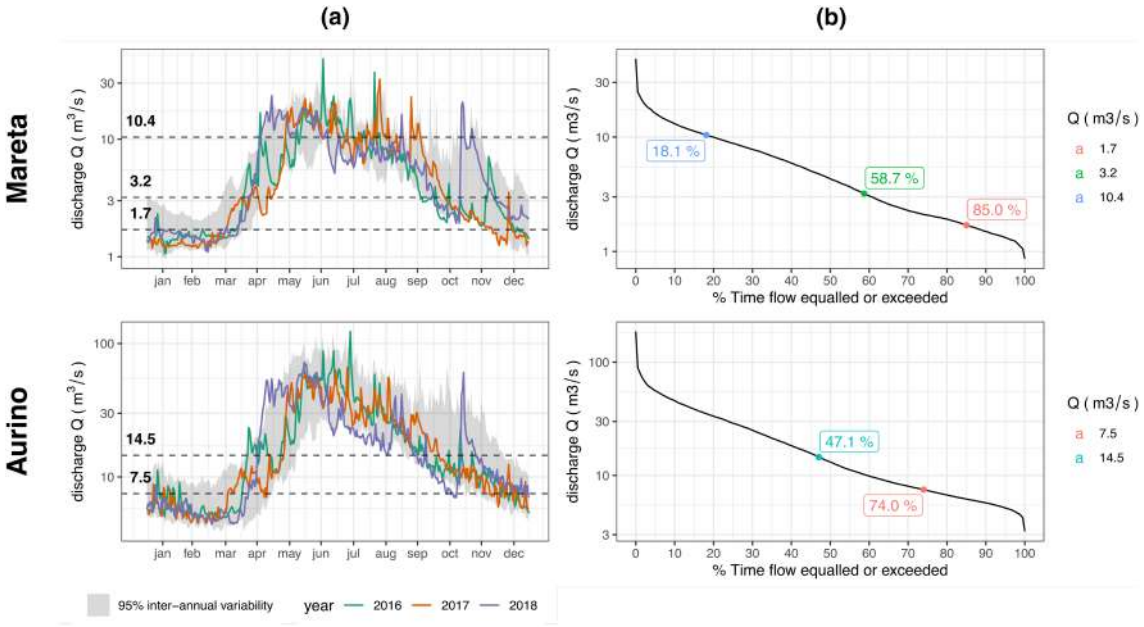

FIGURE 4 Hydrological characteristics of the study reaches on the Mareta (hydrometric station 36750PG in Vipiteno) and Aurino (hydrometric station 59540PG in S. Giorgio) Rivers. Panel (a) depicts yearly hydrographs for the years of surveys (2016–2018), while on the background the 95% variability band is shown (for the period 1986–2020). The dotted horizontal lines visualise the discharges in which the HMU mosaics were surveyed. Panel (b) shows flow duration curves and the surveyed discharges highlighted with the percentage of exceeded time flow

TABLE 2 Description of main hydromorphological units, based on the classification of Vezza et al. (2017) and Belletti et al. (2017)

HMU	Description of hydromorphological characteristics
Backwater	Slack areas along channel margins, caused by eddies behind obstructions, or as part of a side channel. Characterised by shallow and slow-flowing areas
Rapid	Stretches of fast flowing water in high gradient reaches, formed by boulders and large cobbles. Flow is turbulent and characterised by higher air concentration (white-water) or broken standing waves (during low flows). Rapids are characterised by a convex streambed shape.
Riffle	Characterised by shallow and fast flow and uniform sediment which rarely protrudes out of the flow. The flow surface is undulating but unbroken.
Step	Near-vertical drops in the channel bed spanning the entire width. The flow is dominated by spill resistance.
Glide	Characterised by a regular longitudinal bed profile, with a smooth or rippled water surface, flowing parallel to the stream bed.
Pool	Defined by a channel-spanning depression in the channel bed. Pools are characterised by deep and relatively slow velocity flows.

transects of the reach (measured every 10 m), by keeping a steady inflow discharge as upper boundary condition. More information on mesh generation, model set-up and its calibration can be found in Baumgartner (2020).

2.2.4 | The simplified mesohabitat suitability model

Habitat suitability was estimated by means of simplified HSMs for the considered species and life stages. Simplified suitability criteria were based on water depth and flow velocity only, since the focus of this study was the comparison of the suitability of the mesohabitat mosaics predicted by the standard MesoHABSIM approach and the habitat suitability of mesohabitats defined by the segmentation procedure.

The implemented simplified models were constructed based on literature-based information and field observations (Adamczyk et al., 2019; Negro et al., 2021). In particular, the simplified models were built by defining the threshold values of the hydraulic descriptors that maximise the classification performance in terms of presence/absence with the available data. Habitat suitability for the reach (in m^2) was then computed by adding the areas of all suitable mesohabitats, with suitability defined when the model indicates probability of presence.

Models were developed for the following target species (and life stages), selected as representative of the chosen reaches, for the Mareta River, adult and juvenile marble trout (*Salmo marmoratus*) and adult European bullhead (*Cottus gobio*). For the Aurino River, suitability was compared additionally for adult and juvenile grayling (*Thymallus thymallus*).

Examples of the categorical mesohabitat suitability models for the adult and juvenile grayling can be seen in Figure 5. Model description for the other species and life stages can be found in the supporting information.

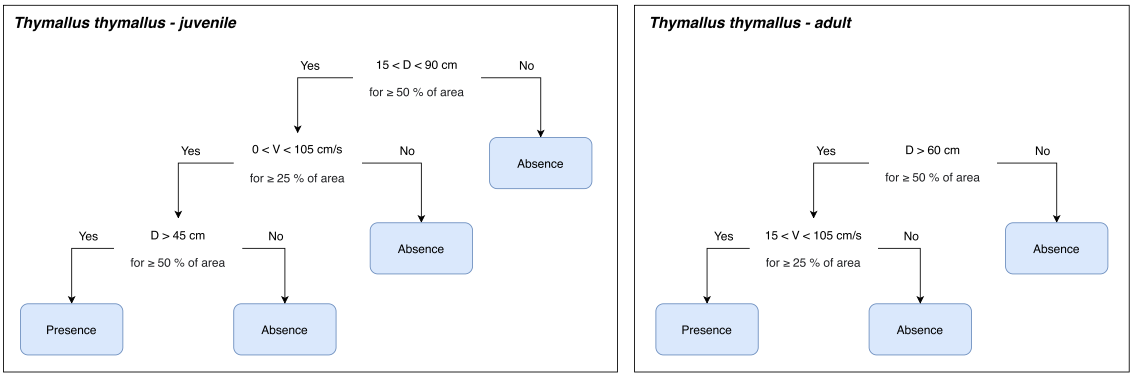

FIGURE 5 Categorical presence/absence biological models, for the juvenile and adult grayling (*Thymallus thymallus*). Variables used are water depth (D) and flow velocity (V), expressed as % of unit area falling in the specified variable range

Note that in the MesoHABSIM approach, mesohabitats are evaluated using a broader range of habitat descriptors (water depth, flow velocity, substrate types and cover availability). Habitat evaluation including only hydraulic descriptors was carried out with the aim of comparing results from the standard and the proposed segmentation procedures.

2.2.5 | Application of the segmentation workflow on the case studies

Clusterisation in step 1 was run for k ranging from 4 to 18. The flow simulated on the computational mesh of the hydraulic model was rasterised to a grid with square cells of 50 cm size (step 2) that was equivalent to the node distance of the computation mesh. For each clusterisation k, the range of segmentation steps from $n = 5$ to 40 segments was evaluated in step 4. The metrics weighted average variance (v), Moran's I (MI), mesonumber (mn) and the resulting GS were computed and optimal segmentations chosen as the combination of k (cluster number) and n(segment number) which minimised the GS. Habitat suitability was then estimated on this optimal mosaic of units.

2.2.6 | Comparison between model-predicted and survey-based mesohabitats

The validity of the proposed mesohabitat modelling methodology was evaluated by comparing reach-scale presence/absence habitat suitability estimates of target fish species from the resulting mesohabitat mosaics with estimates based on ground-survey mapping of HMUs. Resulting habitat-streamflow rating curves were compared for the two reaches of the Mareta and Aurino Rivers. The comparison was performed on the following: each individual curve (for each reach, species and life stage), by grouping all curves from each reach and by grouping all curves from both reaches. The coefficient of determination R^2 and the mean absolute percentage error (MEPA) were computed to assess the goodness of fit between the model- and the survey-based habitat suitability estimations. Further, a qualitative comparison of model- and survey-based habitat-streamflow rating curves was performed, and shape and key features of the curves were compared.

To ensure that differences in covered reach area did not affect results and to render the results comparable, the modelled reach was cropped to the same spatial extension as the surveyed mesohabitat mosaic.

3 | RESULTS

The application of the proposed methodology is presented referring to the case studies of the Aurino and Mareta Rivers. Data collected in the field are first used to derive parameters needed for the segmentation workflow. After presenting the main outcomes of the key workflow steps, modelled habitat suitabilities and rating curves are compared against those obtained from the mesohabsim ground surveys.

3.1 | Choice of workflow parameters based on the case studies

Figure 6 shows the surveyed HMU mosaics for the two study reaches. The reaches are characterised by HMU types typical of piedmont streams with single thread and transitional morphologies (rapid, riffle, glide, pool, backwater and step), with side channels activating at higher discharges. Further, overall channel wet area and average wet channel width show a relevant increase with increasing discharge. The channel width-normalized size distributions of the HMUs ('mesosize' c) are in the range 0.17–8.67, with the largest units, relative to the channel width, found in the Mareta at the lowest discharge (Figure 7). Such distributions can be fitted to a log-normal distribution, and their comparison shows that they can be assumed statistically similar (ANOVA after logarithmic correction:

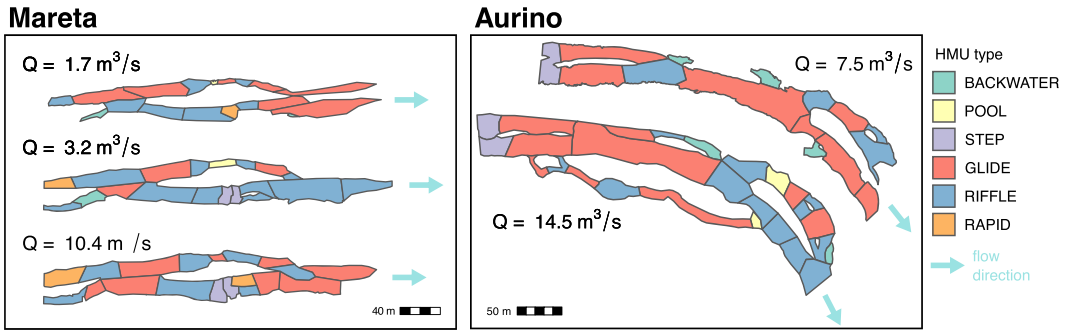

FIGURE 6 Surveyed mesohabitats mosaics for the Mareta and Aurino Rivers

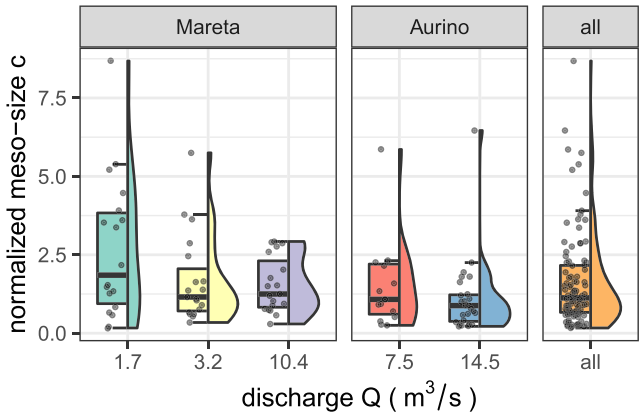

FIGURE 7 Distributions of normalised mesosize values c, for the surveyed HMU mosaics in the Mareta and Aurino Rivers at different discharges. 'All' refers to the distribution of all pooled c values

$F_{(3,73)} = 2.109$, $p = 0.106$). Therefore, a pooled distribution composed by all *c* values from both sites and all five flow conditions was created to compute the parameters required for the segmentation workflow (see also Section 2.1), yielding a logarithmic mean of $\mu_{log} = 0.114$ and a standard deviation of $\sigma_{log} = 0.890$ (Shapiro–Wilk normality test: $W = 0.987$, $p = 0.49$).

In Figure 8, GSs computed for the surveyed mesohabitat mosaics are shown. In this case, the GS is affected only by the average of *v* and *MI*, since the mesonumber would yield the same value for each surveyed mosaic. To allow a comparison between different discharges, rivers and variables, the values of Moran's I have been normalised using the range (−1, 1) as minimum and maximum values and the values of the weighted average variance divided by the overall reach variance. It appears that using flow velocity as the variable to compute the GS consistently yields the lowest score values. The only exception is found in the case of the Mareta reach at the lowest flow $Q = 1.7$ m^3/s, for which the GS computed using the Froude number has a slightly lower value. For the above reasons, water velocity was therefore chosen as the reference variable to compute the modelled GSs from the mesohabitats extracted from the outputs of the hydraulic model.

3.2 | Step-by-step results of the segmentation workflow

Figure 9 shows an example of the flow clusterisation (step 1, Figure 9a) and of the resulting polygonised hydraulic units (step 2, Figure 9b) for the lowest discharge ($Q = 1.7$ m^3/s) at the Mareta River reach. With increasing number of clusters *k* (4, 10 and 18), the resulting bivariate clusters in the velocity-depth parameter space are smaller, resulting in increasingly fragmented mosaics of hydraulic units. While at $k = 4$, a relative low number of very large units forms ($\bar{c} = 0.07$, $c_{max} = 4.15$), often already with sizes close to the mesoscale, at $k = 18$, the overall unit size is very small ($\bar{c} = 0.02$, $c_{max} = 1.45$).

After clusterisation and regionalisation of the initially segmented patches, the region growing algorithm was applied. Examples of the resulting segmentation, which is dependent on the initial clusterisation, can be seen in Figure 9c. Following an initial segmentation with $k = 10$, the resulting mosaics with an overall number *n* of 10, 20 and 40 final regions are shown. Large seed units persist over a wider range of region growing steps, since initially large regions will not be further split; conversely, areas characterised by high fragmentation will merge into uniformly distributed units and will therefore be less affected by the initial segmentation.

Once all possible segmentations were run for the initially chosen range of cluster numbers, GSs for every combination of *n* and *k* were computed for every river reach and each simulated discharge. Figure 10 shows an example of the resulting GS matrix for the Mareta reach at low flow ($Q = 1.7$ m^3/s). While the weighted-area variance *v* decreases for smaller numbers of final segments, Moran's I increases for small number of segments. The mesonumber follows a log-normal distribution which favours the selection of a number of segments in the range 15–30. The final GS is minimised for the combination of $k_{opt} = 12$ and $n_{opt} = 21$.

From the mesohabitat mosaic extracted at every combination of *k* and *n*, it is possible to obtain the corresponding habitat suitability for the entire examined reach at the given discharge value by applying the simplified mesohabitat model described in Section 2.2.4. Figure 11 shows 2 examples of predicted habitat suitability matrices for the adult and juvenile marble trout. The upper row in both matrices above the horizontal black line represents the estimated habitat

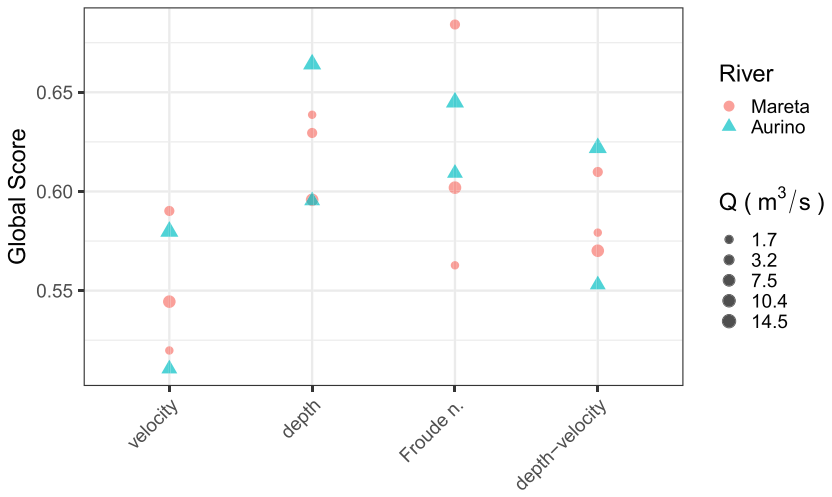

FIGURE 8 Comparison of Global Score values computed on surveyed mesohabitats based on different variables: depth, velocity, Froude number and combined depth and velocity. Global Score values are computed for the two study reaches, for all surveyed discharges ($Q = 1.7$ m^3/s, $Q = 3.2$ m^3/s and $Q = 10.4$ m^3/s for the Mareta, and $Q = 7.5$ m^3/s and $Q = 14.5$ m^3/s for the Aurino River)

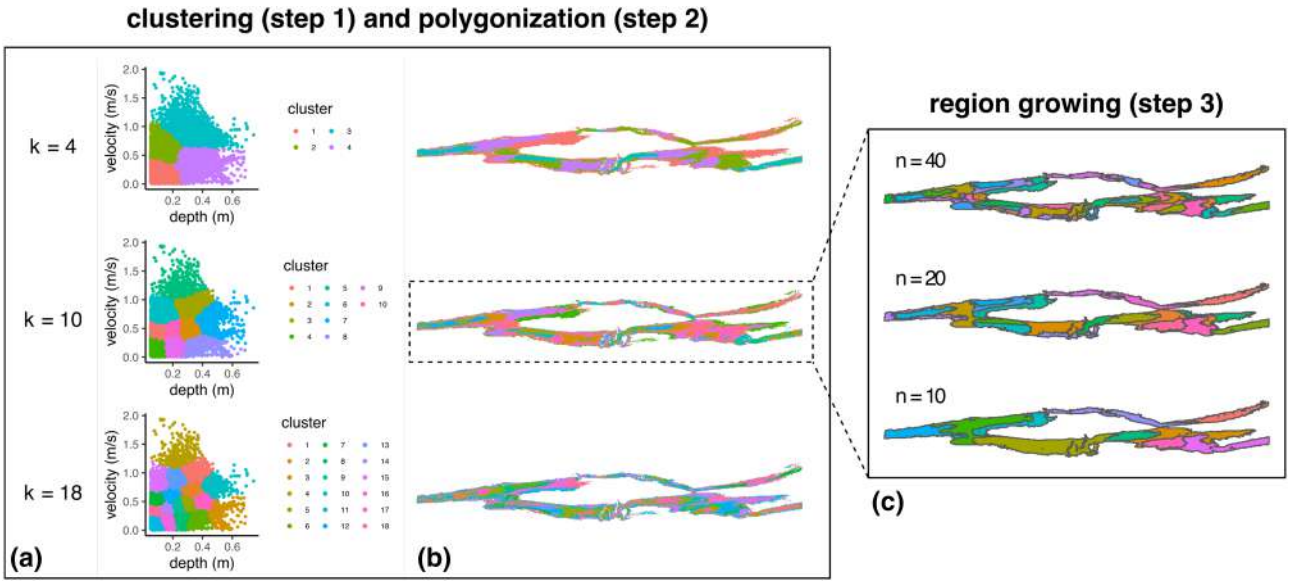

FIGURE 9 Examples of segmentation steps 1–3 based on the Mareta River reach at low flow ($Q = 1.7$ m^3/s). (a) Examples of clusterisation (step 1) and (b) resulting regionalisation (step 2) for different choices of number of clusters k (4, 10, 18). (c) Examples of segmentation after the region growing algorithm is applied (step 3) for an initial clusterisation $k = 10$ and final segmentation numbers n of 10, 20 and 40. Random colours are used to visualise the units

suitability based on the initial segmentation resulting after Step 2. This allows to visualise how the initial segmentation affects the further iterations of the region growing algorithm in terms of habitat suitability. When low k values were used, large initial hydraulic units formed, which cannot be split in successive iterations, and the overall habitat suitability remained more or less constant throughout all iterations, only changing when the number of segments n becomes very low, and the corresponding regions largely increased in size. When the procedure was initialised by larger k values, the initially forming units were smaller, and produced a milder effect on the final suitability. Based on these results, to diminish the initial clusterisation imprint on final patches and therefore on habitat suitabilities, we chose therefore to restrict the k-range for the choice of the optimal combination between 8 and 18.

3.3 | Comparison of habitat-flow rating curves

The segmentation procedure was applied for both reaches and all flow conditions; optimal segmentation combinations of k and n yielding the minimum GS in each case are summarised in Table 3. The corresponding maps can be observed in Figure 12, in which a comparison with the surveyed HMU mosaics is shown.

An example of a spatially explicit comparison between modelled and surveyed habitat suitability maps for the adult European Bullhead in the Mareta reach can be seen in Figure 13. Analogous comparison maps for the other considered species can be seen in the supporting information.

To assess the consistency of the results of the proposed workflow with commonly employed field survey approaches to assess

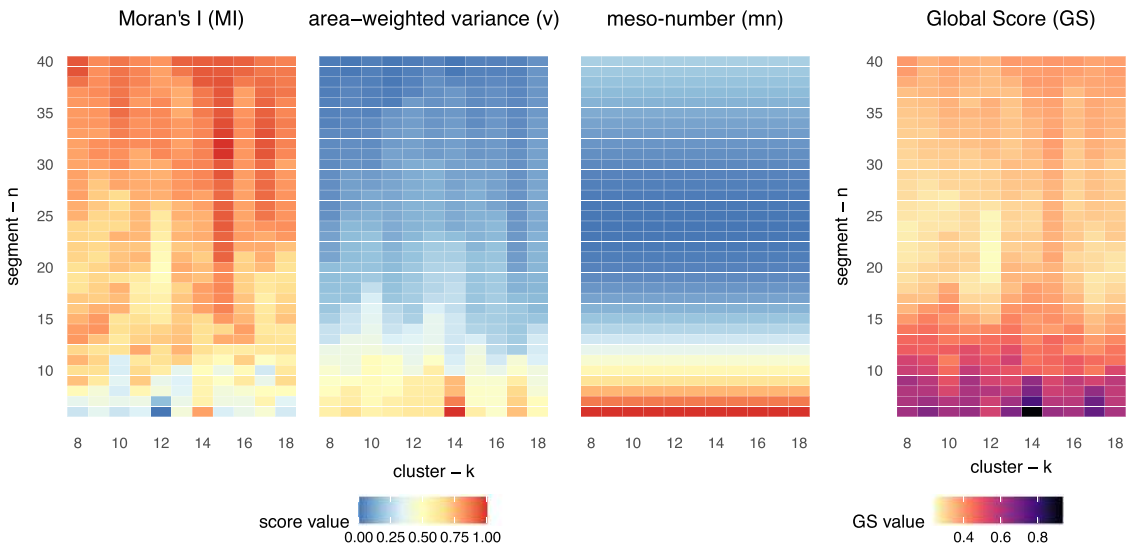

FIGURE 10 Examples of the computed scores (*v*, *MI* and *mn*) and final Global Scores are shown, for the ranges of $k = 8 - 18$ and $n = 6 - 40$. A different colour palette is used for the Global Score, to better highlight the differences in values. The scores were computed for the Mareta reach at low flow ($Q = 1.7$ m^3/s)

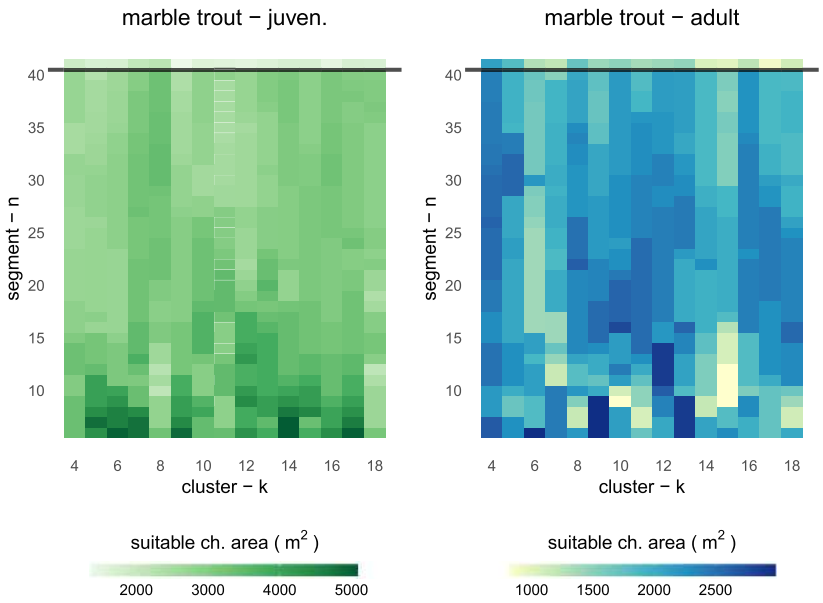

FIGURE 11 Examples of estimated habitat suitability for each combination of *k* and *n* for the juvenile (left) and adult (right) marble trout (in m^2). Results are shown for the ranges $k = 4$–18 and $n = 6$–40. The coloured tiles above the horizontal black line in the graph represent the estimated suitability based on the initial segmentation resulting in step 2. Mareta River at low flow ($Q = 1.7$ m^3/s). Two different colour palettes were used, because of the different ranges of available habitat between the two life stages

mesohabitat suitability, habitat-streamflow rating curves for selected, locally relevant fish species and life stages on two reaches of the Mareta and the Aurino Rivers have been modelled and compared with the ones resulting from the surveyed mesohabitats (Figure 14). A very good overall agreement ($R^2 = 0.91$, MAPE = 27.6 %) between model and survey-based estimates of suitable channel area for both reaches and all species and life stages was found (Figure 15 and Table 4). When comparing individual habitat-streamflow rating curves, the curve with the lowest agreement corresponded to the juvenile marble trout in the Mareta reach ($R^2 = 0.76$, MAPE = 53.2 %), while all other curves had R^2 values above 0.9. Suitability estimates of the modelled

TABLE 3 Selection of the optimal parameters (number of clusters) *k* and (number of segments) *n* for the modelled units

River	Q (m^3/s)	Optimal *k*	Optimal *n*
Mareta	1.7	12	21
	3.2	12	11
	10.4	17	16
Aurino	7.5	14	16
	14.5	14	22

Note: The segmentation procedure was applied in two reaches from the Mareta and Aurino rivers.

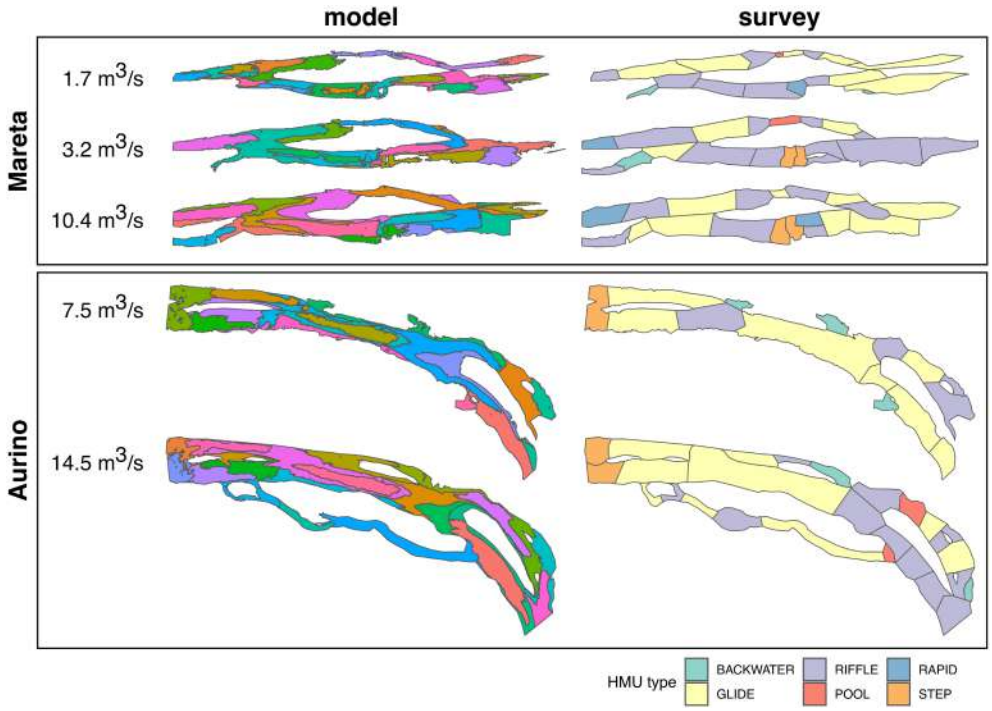

FIGURE 12 Comparison between the modelled (left) and the surveyed (right) mesohabitat mosaic for the Mareta reach and the Aurino reach. On the left, random colours are used to visualise the modelled units, while on the right, the HMUs are represented

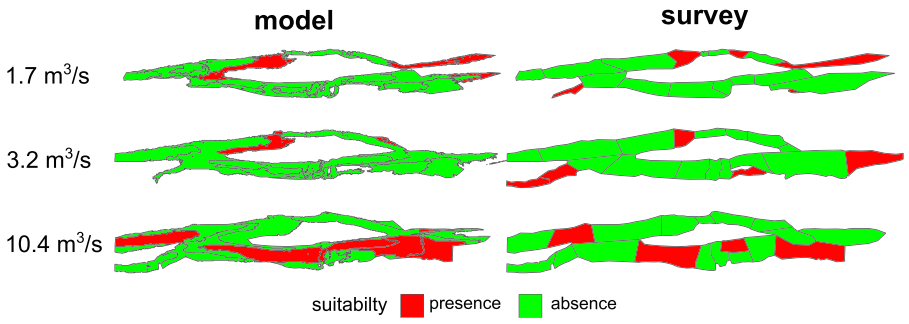

FIGURE 13 Comparison between modelled (left) and surveyed (right) habitat suitability maps for the adult European bullhead at varying discharges in the Mareta reach. Green regions indicate probability of presence and red regions probability of absence

mesohabitats are able to reproduce the magnitude and shape of the curves. As it can be observed, depending on the species flow requirements, the habitat-streamflow rating curve can take different shapes. Some reach a peak and then decrease afterwards for higher discharge values (e.g. for the juvenile marble trout), while in other cases, the curve grows monotonically, as seen for the grayling and the adult marble trout. The modelled curves are able to reproduce both these characteristics. The modelled curve for, for example, the adult bullhead in the Mareta reach well reproduces the magnitude of the estimated suitability over all flow ranges, also reaching a peak at the discharge of 3.2 m^3/s and decreasing at the highest discharge. Although the adult and juvenile life stages of the marble trout have contrasting habitat requirements, resulting in very different habitat-streamflow rating curves, the shape of the curves is quite well reproduced in both river reaches.

4 | DISCUSSION

Application of mesoscale habitat models is increasingly common for a variety of purposes, from ecological flow design, impact assessment and conservation programmes (Wegscheider et al., 2020). Integration with 2D hydraulic modelling allows broader applicability of mesoscale habitat models, extending applications to larger streams and non-wadable flow conditions (Hauer et al., 2009; Wyrick et al., 2014), when on-the-ground and in-stream surveys are challenging or even prohibitive. However, current approaches of algorithm-based mesohabitat modelling require either the definition of thresholds of hydraulic parameters, such as depth, velocity, Froude number or shear stress (e.g. Jowett, 1993; Hauer et al., 2009; Wyrick et al., 2014) or the application of unsupervised clustering techniques (e.g. Legleiter & Goodchild, 2005; Tamminga & Eaton, 2018; van Rooijen et al., 2021). While the definition of thresholds is river-dependent, stage-dependent and requires expert knowledge and ground surveys for calibration of the results (Hauer et al., 2009; Wyrick et al., 2014), the implementation of unsupervised clustering techniques might overcome some of these needs, but the results are still strongly dependent on the choice of clustering algorithm, on data quality and particularly on the choice of the number of classes used for clustering (Legleiter & Goodchild, 2005; Tamminga & Eaton, 2018). Threshold- and clustering-based approaches might result in an excessive fragmentation of units, particularly in complex morphologies and heterogeneous

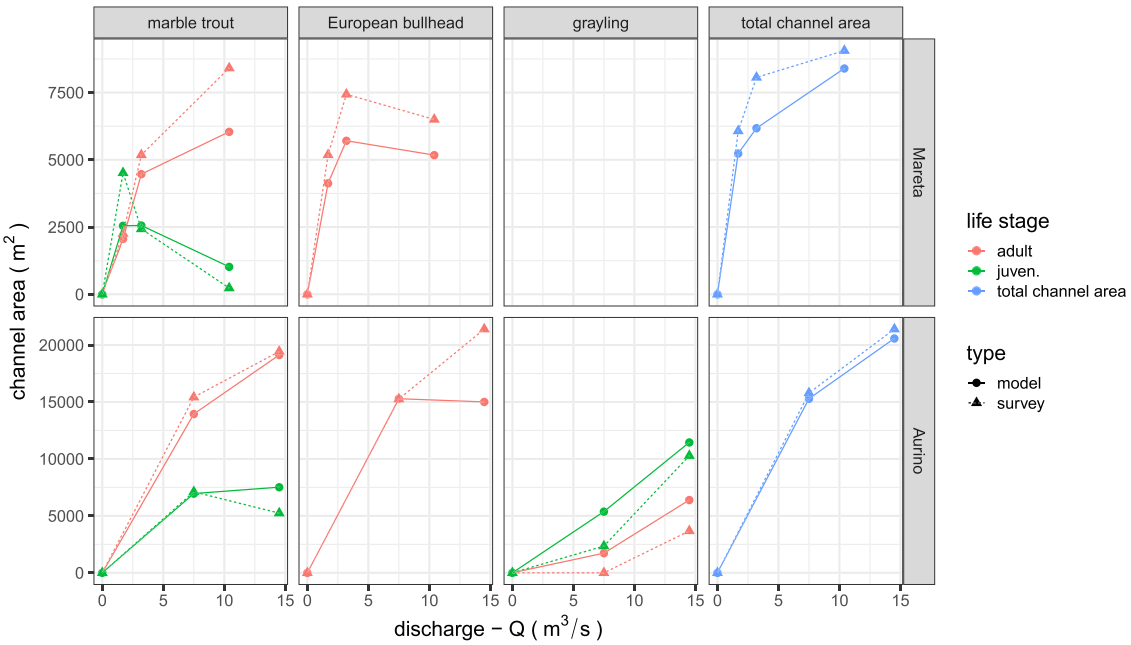

FIGURE 14 Comparison of habitat–streamflow rating curves for the two study reaches in the Mareta and Aurino Rivers, for the following species/life stages: marble trout (adult and juven.) and European bullhead (adult) for Mareta and Aurino; grayling (adult and juven.) only for the Aurino. Points (0, 0) have been added to each curve in this graph, under the assumption that the reach would be dry with vanishing discharge

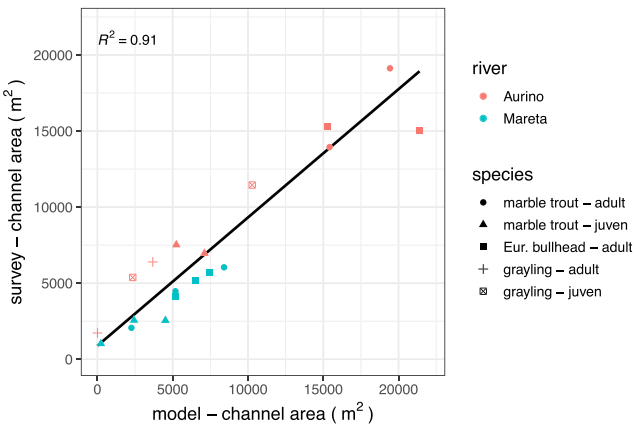

FIGURE 15 Comparison of surveyed and modelled suitable channel areas. The regression line and the coefficient of determination (R^2) are represented

TABLE 4 Comparison of goodness-of-fit statistics for the modelled habitat-streamflow rating curves against survey-based estimates

	Mareta		Aurino	
	R^2	MAPE (%)	R^2	MAPE (%)
Marble trout - adult	0.98	21.5	1.00	6.1
Marble trout - juven.	0.76	53.2	1.00	16.2
Eur. bullhead - adult	0.99	27.1	1.00	21.4
Grayling - adult			1.00	71.2
Grayling -juven			1.00	33.3
All species	0.92	30	0.93	25.3
All species + all rivers	0.91	27.6		

Note: The comparison was done in terms of the coefficient of determination (R^2) and the mean absolute percentage error (MAPE). In 'all species' all points from a reach were grouped, while in 'all species + all rivers' (Figure 15), all points were grouped. Null values (0, 0) as seen in Figure 14 have been excluded from the computation.

flows that is not consistent with a mesoscale ecological approach and river habitat evaluation (Wolter et al., 2016). Implementing spatial contiguity constraints can improve the automated delineation of river mesohabitat patches, creating smoother distributions of patch characteristics and less fragmented patches (van Rooijen et al., 2021). However, while improving on clustering-based approaches, the application by van Rooijen et al. still requires defining the number of final segments characterising the habitat mosaic of the reach.

The methodology presented here builds up on existing approaches, improving on some of their limitations. To avoid the need of defining river-dependent and expert knowledge-based hydraulic thresholds, the delineation of the mesohabitat mosaic is entirely unsupervised, from the automatic delineation of mesohabitats to the selection of segmentation parameters. While the results of the mesohabitat delineation are still dependent on the quality of the modelled flow field and on the choice of the implemented algorithms in step 1 (k-means clustering) and step 3 (region growing), the segmentation parameters, and thus the final segmentation, will be objectively chosen based on an optimality criteria, as defined by the GS (step 4). This ensures that within the range of possible segmentations, only the one that minimises an objectively defined GS is chosen as optimal. Compared with parametric classification approaches, for which parameter

thresholds need to be defined for hydraulic variables such as Froude number, velocity and depth (Hauer et al., 2009; Wyrick et al., 2014), no calibration or expert knowledge is required in the presented approach. To compute the GS, the only parameter that needs to be defined empirically is the mesonumber, while the area-weighted variance and Moran's I are computed based on intrinsical spatial properties of the regions (Espindola et al., 2006). Computation of the empirically defined mesosize logarithmic distribution parameters is also objective and requires only surveyed maps of mesohabitats from the same reach or reaches which are morphologically similar. This ensures that the entire process of segmentation is fully unsupervised, requiring no subjective choice by the operator in defining hydraulic thresholds or segmentation parameters. Moreover, the implemented region growing algorithm (step 3) ensures that the final size of segments is consistent with an empirically defined river mesoscale, avoiding unnecessary fragmentation of hydraulic units, even in complex morphologies with highly heterogeneous flows.

The validity of the methodology was assessed by comparing the results with mesohabitat mosaics mapped according to the Meso-HABSIM methodology, in terms of presence/absence habitat suitability estimates for selected fish target species and life stages. Habitat suitability was assessed for adult and juvenile marble trout, adult European bullhead and adult and juvenile grayling. Modelled results at the reach-scale agree consistently well with survey-based habitat suitability estimates. Overall R^2 values between model- and survey-based results for all analysed species and life stages, and in both rivers, are above 0.9 (with the exception of the juvenile grayling curve with a value of 0.76), showing therefore the potential applicability of the methodology for habitat modelling at the mesoscale.

The model tended to underestimate available suitable habitats in the Mareta reach, while it overestimated their extent in the Aurino reach. Some of the observed differences can be explained by differences between surveyed and modelled wetted channel area, which were particularly pronounced in the Mareta at higher flows (Figure 14). Differences in wetted channel area could be caused by mapping uncertainties (Poole et al., 1997), which were performed by using a handheld rtk-GPS or rangefinder and roughly mapping the shape of the visually observed units. This could result in an overestimation of HMU area due to irregular shapes or presence of large boulders within the unit, which would add to the surveyed wet area while remaining dry in the hydraulic model results. Furthermore, while very shallow marginal channel areas have been included in the HMUs during field surveys, a 5 cm threshold was applied to the 2D hydraulic modelling results, which further diminished channel extent in marginal shallow areas. It must also be noted that while transitions between HMUs are fuzzy, the mapped mesohabitats are discrete, and hence, the transitional area between units has to be subjectively assigned by the operator to one unit or the other while surveying in the field. The segmentation performed by the region growing algorithm is not constrained by the same mapping limitations of field surveys and is therefore better suited to render smooth transitions between units. This can be observed in the differences in shapes between mapped and modelled mesohabitats (Figure 12). The resulting units might therefore better represent areas within the channel characterised by homogeneous hydraulic conditions, and having distinct hydraulic characteristics than neighbouring units (e.g. Figure S11 in supporting information), compared with on-the-ground field surveys. The methodology allows hence an objective and repeatable segmentation of the reach into a mosaic of hydromorphologically based units at a given discharge. Resulting habitat suitability maps are computed using HSMs defined at a scale that is consistent with the scale of the modelled units. Not surprisingly, there is a high degree of overlap of suitable and unsuitable areas between modelled and surveyed units (see, e.g., Figure 13 and Section S4 in the supporting information for further analysis and examples). Resulting suitability maps can potentially better describe suitable and unsuitable reach areas, due to the model's ability to objectively differentiate hydromorphological conditions within the reach (by, e.g., aggregating longer shallower bank areas together), and allowing for smoother transitions between contiguous units (by, e.g., separating midchannel areas characterised by faster and deeper currents from lateral slower flowing areas). It must be noted, however, that mesohabitat segmentation was done using only hydraulic parameters, while HMUs are recognised in the field also considering a wider range of hydromorphological characteristics (Belletti et al., 2017), which might further explain differences in shape between modelled and surveyed mesohabitats.

The two reaches in the Mareta and Aurino Rivers, although having different hydromorphological characteristics, share similar mesoscale size distributions (Figure 7). River systems are shaped across a multitude of spatial and temporal scales (Belletti et al., 2017; Frissell et al., 1986), which are hierarchically structured, with processes and forms occurring at the lower scales being dependent on the morphological dynamics of the higher scales and on the environmental boundary conditions (e.g. valley slope, sediment size and riparian vegetation) posed by the lower scales. Geomorphic units or HMUs in particular scale with the channel width (Belletti et al., 2017; Frissell et al., 1986; Gurnell et al., 2016). This might not be surprising, as many river channel forms have close association with channel geometry features such as channel size. Examples include the length scale of alternate bars (Adami et al., 2016; Tubino et al., 1999) and of meander bends (Leopold & Wolman, 1960; Vermeulen et al., 2016). Further research is needed to assess to which extent the characteristic scale range observed in this study is universal or river and morphology dependent.

While one of the key advantages of the proposed methodology is its ability to segment the flow of a river reach into mesoscale habitat patterns through intrinsical properties of the arising regions, no real classification of these regions into HMU types can be made. Classifying these regions into, for example, riffles and pools would require a posteriori classification by the operator. These makes the proposed methodology inappropriate to study, for example, temporal and spatial dynamics of HMUs, since no information on the amount and distribution of specific HMU types can be extracted. Further, segmentation of the reach was based only on the hydraulic parameters water depth and flow velocity, neglecting other hydrological and morphological variables which are also commonly used to classify and distinguish between mesohabitats, such as water surface elevation gradient, surface flow

patterns, substrate character and channel form (Belletti et al., 2017; Borsányi et al., 2004; Frissell et al., 1986; Gurnell et al., 2016; Parasiewicz, 2007; Wegscheider et al., 2020). Although an optimal segmentation can be objectively identified based on the GS, the methodology does not ensure that a global optimal solution is found, since only a combination of segmentation options arising from different initial clusterisations and final number of segments is analysed and not the entire parameter space range. To limit this issue, the proposed methodology is initialised multiple times, by setting a wider range of number of clusters k (step 1), which allows creating different initial seed segmentations (step 2). Future model developments could improve current methodological limitations, by integrating more variables into the segmentation process and testing different segmentation algorithms (e.g. Anselin, 2005; Grubesic et al., 2014). In regard to habitat suitability modelling presented in this work, only hydraulic parameters were used, specifically flow depth and velocity. More comprehensive mesohabitat models also require information on substrate-type distributions, presence/absence of covers and others (Parasiewicz, 2007; Vezza et al., 2014). These parameters need to be integrated using other approaches, for example, by surveys or remote sensing techniques.

Finally, a key methodological novelty of the presented approach is the adaptation of methods commonly used in image segmentation to the field of river ecohydraulics, which, however, do not seem to have been used for ecohydraulics applications so far. In the ecological field, image segmentation and segmentation quality assessment have been, for example, implemented to map and classify sensitive marine habitats (Ventura et al., 2018), boreal forest habitats (Räsänen et al., 2013) and semi-urban landscapes (Kavzoglu et al., 2017). While the first two indexes used to compute the GS, Moran's I and the weighted averaged variance, are derived from the field of object-based image segmentation (e.g. Espindola et al., 2006; Kavzoglu et al., 2017; Räsänen et al., 2013; Ventura et al., 2018), the mesonumber was introduced for the purpose of mesoscale habitat segmentation. This index ensures that the final segmentation results are consistent in size with the empirically defined mesoscale. Such scale will likely vary depending on river morphology and mesohabitat classification system used (Frissell et al., 1986; Gurnell et al., 2016; Rinaldi et al., 2013). While having been developed and tested on gravel-bed rivers, the methodology is flexible and has the potential to be adapted to suit the needs of the user and be of value also for a wider range of river types or mesoscale classification systems used (Belletti et al., 2017; Borsányi et al., 2004; Frissell et al., 1986; Gurnell et al., 2016; Hauer et al., 2009; Parasiewicz, 2007; Wegscheider et al., 2020; Wyrick et al., 2014). In particular, these can be achieved through the implementation of the empirically defined mesosize distribution, which defines the scale of the resulting modelled units.

5 | CONCLUSIONS

This study has shown the implementation of an unsupervised workflow to segment the spatial distribution of flow depth and velocity obtained from two-dimensional depth-averaged hydraulic modelling applied to quantify mesoscale habitat suitability in a river reach. An optimal segmentation is defined objectively based on minimising a GS, obtained as the average of three metrics that represent (i) intrasegment homogeneity, (ii) intersegment heterogeneity of the regions and (iii) an empirically based optimal number of segments consistent with the mesoscale, that is, the scale of the so-called hydromorphological units. The workflow is river-independent and fully unsupervised, as it does not require calibration or subjective choices of segmentation parameters. The validity of the methodology to reconstruct habitat-streamflow rating curves that are at least equivalent to those obtained from survey-based HMU mosaics has been shown for two study reaches at morphologically distinct rivers, the wandering Mareta and the meandering Aurino. An overall agreement of $R^2 = 0.91$ is found for all analysed species and life stages. The workflow can be applied in the context of large rivers and nonwadable flow conditions and has the key advantage of broadening the applicability of habitat models at the mesoscale.

ACKNOWLEDGEMENTS

Financial support was received by the Italian Ministry of Education, University and Research (MIUR) under the Departments of Excellence, grant L.232/2016 and by the Autonomous Province of Bolzano/Bozen, under the research project 'FHARMOR Fish Habitat in Alpine rivers: integrating monitoring, modelling and remote sensing', funding agreement no. 17/34 of 03/11/2016. We also thank the reviewers for their helpful comments on the manuscript. Open Access Funding provided by Universita degli Studi di Trento within the CRUI-CARE Agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

David Farò https://orcid.org/0000-0003-3168-8162
Katharina Baumgartner https://orcid.org/0000-0002-8888-882X
Paolo Vezza https://orcid.org/0000-0002-6784-8036
Guido Zolezzi https://orcid.org/0000-0003-2807-7387

REFERENCES

Adamczyk, M., Parasiewicz, P., Vezza, P., Prus, P., & De Cesare, G. (2019). Empirical validation of mesoHABSIM models developed with different habitat suitability criteria for Bullhead Cottus Gobio L. as an indicator species. *Water*, *11*(4), 726.

Adami, L., Bertoldi, W., & Zolezzi, G. (2016). Multidecadal dynamics of alternate bars in the Alpine Rhine River. *Water Resources Research*, *52*(11), 8938–8955.

Ahmadi-Nedushan, B., St-Hilaire, A., Bérubé, M., Robichaud, E., Thiémonge, N., & Bobée, B. (2006). A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications*, *22*(5), 503–523.

Anselin, L. (2005). Exploring spatial data with GeoDa: A workbook. *Geography*, 244.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256.

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the annual acm-siam symposium on discrete algorithms*.

Baumgartner, K. (2020). Analyse und Evaluierung der praktischen Anwendung von topo-bathymetrischen LiDAR Daten in alpinen Gewässern. (Ph.D. Thesis).

Belletti, B., Rinaldi, M., Bussettini, M., Comiti, F., Gurnell, A. M., Mao, L., Nardi, L., & Vezza, P. (2017). Characterising physical habitats and fluvial hydromorphology: A new system for the survey and classification of river geomorphic units. *Geomorphology*, *283*, 143–157.

Bisson, P. A., Nielsen, J. L., Palmason, R. A., & Grove, L. E. (1982). A system of naming habitat types in small streams, with examples of habitat utilization by salmonids during low stream flow. In *Acquisition and utilization of aquatic habitat inventory information*.

Böck, S., Immitzer, M., & Atzberger, C. (2017). On the Objectivity of the Objective FunctionâÄŤProblems with Unsupervised Segmentation Evaluation Based on Global Score and a Possible Remedy. *Remote Sensing*, *9*(8), 769.

Borsányi, P., Alfredsen, K., Harby, A., Ugedal, O., & Kraxner, C. (2004). A meso-scale habitat classification method for production modelling of Atlantic Salmon in Norway. *Hydroécologie Appliquée*, *14*, 119–138.

Bovee, K. D. (1982). *A Guide to Stream Habitat Analysis Using the Instream Flow Incremental Methodology*: US Fish and Wildlife Service FWS/OBS.

Brierley, G. J., & Fryirs, K. (2000). River styles, a geomorphic approach to catchment characterization: Implications for river rehabilitation in Bega Catchment, New South Wales, Australia. *Environmental Management*, *25*(6), 661–679.

Campana, D., Marchese, E., Theule, J. I., & Comiti, F. (2014). Channel degradation and restoration of an Alpine river and related morphological changes. *Geomorphology*, *221*, 230.

Clifford, N. J., Harmar, O. P., Harvey, G., & Petts, G. E. (2006). Physical habitat, eco-hydraulics and river design: A review and re-evaluation of some popular concepts and methods. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *16*(4), 389–408.

Davis, J. A., & Barmuta, L. A. (1989). An ecologically useful classification of mean and near-bed flows in streams and rivers. *Freshwater Biology*, *21*(2), 271–282.

Demarchi, L., Bizzi, S., & Piégay, H. (2016). Hierarchical object-based mapping of riverscape units and in-stream mesohabitats using LiDAR and VHR imagery. *Remote Sensing*, *8*(2), 97.

Eisner, A., Young, C., Schneider, M., & Kopecki, I. (2005). MesoCASiMiR—New mapping method and comparison with other current approaches.

Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., & Monteiro, A. M. (2006). Parameter selection for regionâĂŘgrowing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, *27*(14), 3035–3040.

Farò, D., Andreoli, A., Aufleger, M., Baran, R., Baumgartner, K., Bertoldi, W., Bussettini, M., Carolli, M., Comiti, F., Demarchi, L., Jocham, S., Klar, R., Marangoni, N., Parasiewicz, P., Politti, E., Scorpio, V., Steinbacher, F., Vezza, P., & Zolezzi, G. (2018). FHARMOR: Fish Habitat in Alpine Rivers - Integrating Monitoring, Modelling and Remote sensing. In *I.s. rivers 2018 - 3rd international conference on integrative sciences and sustainable development of rivers*, pp. 221.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*: Sage.

Frissell, C. A., Liss, W. J., Warren, C. E., & Hurley, M. D. (1986). A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management*, *10*(2), 199–214.

Grubesic, T. H., Wei, R., & Murray, A. T. (2014). Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Annals of the Association of American Geographers*, *104*(6), 1134–1156. https://doi.org/10.1080/00045608.2014.958389

Gurnell, A. M., Rinaldi, M., Belletti, B., Bizzi, S., Blamauer, B., Braca, G., Buijse, A. D., Bussettini, M., Camenen, B., Comiti, F., Demarchi, L., García de Jalón, D., González del Tánago, M., Grabowski, R. C., Gunn, I. D. M., Habersack, H., Hendriks, D., Henshaw, A. J., Klösch, M., …, & Ziliani, L. (2016). A multi-scale hierarchical framework for developing understanding of river behaviour to support river management. *Aquatic Sciences*, *78*(1), 1–16.

Harper, D. M., Smith, C. D., & Barham, P. J. (1992). Habitats as the building blocks for river conservation assessment. *River conservation and management*, 311.

Hauer, C., Mandlburger, G., & Habersack, H. (2009). Hydraulically related hydro-morphological units: Description based on a new conceptual mesohabitat evaluation model (MEM) using LiDAR data as geometric input. *River Research and Applications*, *25*(1), 29–47.

Hauer, C., Unfer, G., Tritthart, M., Formann, E., & Habersack, H. (2011). Variability of mesohabitat characteristics in riffle-pool reaches: Testing an integrative evaluation concept (FGC) for MEM-application. *River Research and Applications*, *27*(4), 403–430.

Jowett, I. G. (1993). A method for objectively identifying pool, run, and riffle habitats from physical measurements. *New Zealand Journal of Marine and Freshwater Research*, *27*(2), 241–248.

Kavzoglu, T., Erdemir, M. Y., & Tonbul, H. (2017). Classification of semiurban landscapes from very high-resolution satellite images using a regionalized multiscale segmentation approach. *Journal of Applied Remote Sensing*, *11*(03), 1.

Kemp, J. L., Harper, D. M., & Crosa, G. A. (1999). Use of 'functional habitats' to link ecology with morphology and hydrology in river rehabilitation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *9*(1), 159–178.

Legleiter, C. J., & Goodchild, M. F. (2005). Alternative representations of in-stream habitat: Classification using remote sensing, hydraulic modeling, and fuzzy logic. *International Journal of Geographical Information Science*, *19*(March 2015), 29–50.

Leopold, L. B., & Wolman, M. G. (1960). River meanders. *Geological Society of America Bulletin*, *71*(6), 769–793.

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Moir, H. J., & Pasternack, G. B. (2008). Relationships between mesoscale morphological units, stream hydraulics and Chinook salmon (Oncorhynchus tshawytscha) spawning habitat on the Lower Yuba River, California. *Geomorphology*, *100*(3-4), 527–548.

Negro, G., Fenoglio, S., Quaranta, E., Comoglio, C., Garzia, I., & Vezza, P. (2021). Habitat preferences of Italian freshwater fish: A systematic review of data availability for applications of the MesoHABSIM model. *Frontiers in Environmental Science*, *0*, 305.

Newson, M. D., & Newson, C. L. (2000). Geomorphology, ecology and river channel habitat: Mesoscale approaches to basin-scale challenges. *Progress in Physical Geography: Earth and Environment*, *24*(2), 195–217.

Nujić, M., & Hydrotec (2017). Benutzerhandbuch HYDRO_AS-2D, 2D-Strömungsmodell für die wasserwirtschaftliche Praxis.

Padmore, C. L. (1998). The role of physical biotopes in determining the conservation status and flow requirements of british rivers. *Aquatic Ecosystem Health and Management*, *1*(1), 25–35.

Parasiewicz, P. (2007). The MesoHABSIM model revisited. *River Research and Applications*, *23*(8), 893–903.

Parasiewicz, P., Castelli, E., Rogers, J. N., & Plunkett, E. (2012). Multiplex modeling of physical habitat for endangered freshwater mussels. *Ecological Modelling*, *228*, 66–75.

Poole, G. C., Frissell, C. A., & Ralph, S. C. (1997). In-stream unit classification: Inadequacies for monitoring and some consequences for management. *Journal of the American Water Resources Association*, *33*(4), 879–896.

Räsänen, A., Rusanen, A., Kuitunen, M., & Lensu, A. (2013). What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing*, *34*(23), 8603–8627.

Rinaldi, M., Surian, N., Comiti, F., & Bussettini, M. (2013). A method for the assessment and analysis of the hydromorphological condition of Italian

streams: The Morphological Quality Index (MQI). *Geomorphology*, *180-181*, 96–108.

Rivas Casado, M., Ballesteros González, R., Ortega, J. F., Leinster, P., & Wright, R. (2017). Towards a transferable UAV-based framework for river hydromorphological characterization. *Sensors 2017*, *Vol. 17*, *Page 2210*, *17*(10), 2210.

Rivas Casado, M., Gonzalez, R. B., Kriechbaumer, T., & Veal, A. (2015). Automated identification of river hydromorphological features using UAV high resolution aerial imagery. *Sensors 2015*, *Vol. 15*, *Pages 27969-27989*, *15*(11), 27,969–27,989.

Tamminga, A., & Eaton, B. (2018). Linking geomorphic change due to floods to spatial hydraulic habitat dynamics. *Ecohydrology*, *11*(8), e2018.

Tubino, M., Repetto, R., & Zolezzi, G. (1999). Free bars in rivers. *Journal of Hydraulic Research*, *37*(6), 759–775.

van Rooijen, E., Vanzo, D., Vetsch, D. F., Boes, R. M., & Siviglia, A. (2021). Enhancing an unsupervised clustering algorithm with a spatial contiguity constraint for river habitat analysis. *Ecohydrology*, *14*, e2285.

Ventura, D., Bonifazi, A., Gravina, M. F., Belluscio, A., & Ardizzone, G. (2018). Mapping and classification of ecologically sensitive marine habitats using unmanned aerial vehicle (UAV) imagery and object-based image analysis (OBIA). *Remote Sensing*, *10*(9), 1331. http://www.mdpi.com/2072-4292/10/9/1331

Vermeulen, B., Hoitink, A. J. F., Zolezzi, G., Abad, J. D., & Aalto, R. (2016). Multiscale structure of meanders. *Geophysical Research Letters*, *43*(7), 3288–3297.

Vezza, P., Parasiewicz, P., Spairani, M., & Comoglio, C. (2014). Habitat modeling in high-gradient streams: The mesoscale approach and application. *Ecological Applications*, *24*(4), 844–861.

Vezza, P., Zanin, A., & Parasiewicz, P. (2017). Manuale tecnico-operativo per la modellazione e la valutazione dell'integrità dell'habitat fluviale.

Wadeson, R. A., & Rowntree, K. M. (1998). Application of the hydraulic biotope concept to the classification of instream habitats. *Aquatic Ecosystem Health and Management*, *1*(2), 143–157.

Wallis, C., Maddock, I., Visser, F., & Acreman, M. C. (2012). A framework for evaluating the spatial configuration and temporal dynamics of hydraulic patches. *River Research and Applications*, *28*(5), 585–593.

Wegscheider, B., Linnansaari, T., & Curry, R. A. (2020). Mesohabitat modelling in fish ecology: A global synthesis. *Fish and Fisheries*, *21*, 927.

Wolter, C., Buijse, A. D., & Parasiewicz, P. (2016). Temporal and spatial patterns of fish response to hydromorphological processes. *River Research and Applications*, *32*(2), 190–201.

Woodget, A. S., Visser, F., Maddock, I. P., & Carbonneau, P. E. (2016). The accuracy and reliability of traditional surface flow type mapping: Is it time for a new method of characterizing physical river habitat? *River Research and Applications*, *32*(9), 1902–1914.

Wyrick, J. R., Senter, A. E., & Pasternack, G. B. (2014). Revealing the natural complexity of fluvial morphology through 2D hydrodynamic delineation of river landforms. *Geomorphology*, *210*, 14–22.

Zerbe, S., Scorpio, V., Comiti, F., & Rohrmoser, O. (2019). Vegetationsentwicklung nach einer Flussrenaturierung in den Alpen. WASSERWIRTSCHAFT.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Farò, D., Baumgartner, K., Vezza, P., & Zolezzi, G. (2022). A novel unsupervised method for assessing mesoscale river habitat structure and suitability from 2D hydraulic models in gravel-bed rivers. *Ecohydrology*, *15*(7), e2452. https://doi.org/10.1002/eco.2452