

Scale effects on the performance of niche-based models of freshwater fish distributions

Oskar Kärcher^{a,b,*}, Karin Frank^{c,d,f}, Ariane Walz^b, Danijela Markovic^a

^a Faculty of Business Management and Social Sciences, Osnabrück University of Applied Sciences, Caprivistr. 30A, 49076, Osnabrück, Germany

^b Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476, Potsdam, Germany

^c UFZ – Helmholtz Centre for Environmental Research Ltd, Department for Ecological Modelling, Permoserstr. 15, 04318, Leipzig, Germany

^d Institute of Environmental Systems Research, University of Osnabrück, Barbarastr. 12, 49076, Osnabrück, Germany

^f iDiv – German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig, Deutscher Platz 5e, 04103, Leipzig, Germany

ARTICLE INFO

Keywords:

Catchment order
Conservation planning
Danube
Generalized additive models
Species distribution modelling

ABSTRACT

Niche-based species distribution models (SDMs) have become an essential tool in conservation and restoration planning. Given the current threats to freshwater biodiversity, it is of fundamental importance to address scale effects on the performance of niche-based SDMs of freshwater species' distributions. The scale effects are addressed here in the context of hierarchical catchment ordering, considered as counterpart to coarsening grain-size by increasing grid-cell size. We combine fish occurrence data from the Danube River Basin, the hierarchical catchment ordering and multiple environmental factors representing topographic, climatic and anthropogenic effects to model fish occurrence probability across multiple scales. We focus on 1st to 5th order catchments. The spatial scale (hierarchical catchment order) only marginally influences the mean performance of SDMs, however the uncertainty of the estimates increases with scale. Key predictors and their relative importance are scale and species dependent. Our findings have useful implications for choosing proper species dependent spatial scales for river rehabilitation measures, and for conservation planning in areas where fine grain species data are unavailable.

1. Introduction

Niche-based species distribution models (SDMs), play a central role in studying species habitat preferences, conservation and restoration planning at global, regional and local scales (e.g. Franklin, 2009). Multitudes of different algorithms now exist, and new methods and algorithms are developed continuously (e.g. Efron and Hastie, 2016). Methodological aspects of SDMs have been thoroughly studied from many perspectives, demonstrating reasons for differing performances (see Elith and Graham, 2009) including the sampling patterns (presence/absence or presence only data, Brotons et al., 2004), choice of single algorithm vs. consensus methods (Marmion et al., 2009), the length of studied future time periods and used predictors (Morán-Ordóñez et al., 2016), collinearity (Dormann et al., 2013), prevalence (Lawson et al., 2014) and the choice of grain-size (resolution) of environmental layers (Guisan et al., 2007). While the superior performance of consensus SDM methods has been repeatedly demonstrated increasing the reliability of the projections and reducing the model

uncertainties (e.g. Thuiller, 2004; Marmion et al., 2009; Grenouillet et al., 2011), the influence of spatial scales on predictive power was less frequently studied. A study by Guisan et al. (2007) has shown that a 10-fold increase of the grid cell size does not have a substantial effect on the performance of SDMs describing distributions of bird, plant and vertebrate species. Similarly, Lauzeral et al. (2013) have shown a minor decrease in the performance of SDMs for five virtual species with coarsening grain-size from 30'' × 30'' to 32'' × 32''. Comparable studies that use multiple grain-size resolutions and SDMs for freshwater species do not exist, however, a study by Domisch et al. (2013) has shown that the choice of stream network or landscape as study area does not affect the performance of SDMs of stream biota.

Hierarchical catchment ordering provides a valuable framework to address scale effects on niche-based SDMs of freshwater species' distributions (Allan et al., 1997) and was thus considered as counterpart to coarsening grain-size by increasing grid-cell size, which is commonly used in modelling distributions of virtual or terrestrial species. Specifically, the hierarchical structure of catchments is considered to be one

* Corresponding author at: Faculty of Business Management and Social Sciences, Osnabrück University of Applied Sciences, Caprivistr. 30A, 49076, Osnabrück, Germany.

E-mail address: o.kaercher@hs-osnabrueck.de (O. Kärcher).

<https://doi.org/10.1016/j.ecolmodel.2019.05.006>

Received 10 December 2018; Received in revised form 7 May 2019; Accepted 12 May 2019

Available online 21 May 2019

0304-3800/ © 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1
Initially considered environmental variables.

Category	Variable	Abbreviation	Description
Topographic	Altitude	Altitude	Mean catchment elevation
	Gradient	Gradient	Relief energy of the river segment
	Length	Length	Cumulative length of the upstream flow network
	Slope	Slope	Mean slope in percent
Climatic	Mean temperature	AnnTMean	Mean annual temperature
	Maximum temperature	AnnTMax	Max temperature of warmest month
	April–September temperature	TAprSep	Mean annual April to September temperature
	May–August temperature	TMaiAug	Mean annual May to August temperature
	Temperature range	TRange	Temperature annual range
	Annual temperature seasonality	TSeason	Standard deviation
	Wettest quarter temperature	WettTMean	Mean temperature of wettest quarter
	Precipitation	AnnPMean	Mean annual precipitation
	April–September precipitation	PAprSep	Mean annual April to September precipitation
	May–August precipitation	PMaiAug	Mean annual Mai to August precipitation
	Precipitation seasonality	PSeason	Coefficient of variation
Anthropogenic	Agricultural area	Agriculture	% of catchment area used for agriculture
	RowCrops	RowCrops	% of catchment area under row crops
	Forest	Forest	% of catchment area under forest
	Grassland	Grassland	% of catchment area under grasslands
	Pastures	Pastures	% of catchment area under pastures
	Built-up area	BuiltUp	% of catchment area under artificial surfaces
	Population	Population	Number of inhabitants per catchment

of the key challenges in the application of SDMs in freshwater ecosystems (Domisch et al., 2015). Given that SDMs are central to both fundamental and applied research in biogeography (Araújo and Guisan, 2006), in particular conservation and restoration planning, the need for an improved understanding of the effects of grain-size on the overall performance of freshwater species SDMs becomes apparent.

Studies of freshwater species' distributions using multiple spatial scales are rare and generally limited to the comparison of the relative importance of reach- and sub-catchment-related environmental factors for the variation in species local composition (e.g. Hopkins and Burr, 2009; Esselman and Allan, 2010) using exploratory data analysis techniques. However, Domisch et al. (2013) studied the effects of the extent of the modelled area (stream network vs. landscape) and the choice of predictors on SDMs of stream macroinvertebrates. As the lack of data constrains freshwater conservation strategies in many regions of the world, in particular less-developed countries, catchment scale data hold great promise where fine grain survey data on freshwater species occurrence or environmental factors are unavailable. Our study aims to accommodate the need for clarity on the predictive performance of freshwater SDMs based on catchment scale data, i.e., for SDMs lacking fine grain survey data. Given the current threats to freshwater biodiversity including high anthropogenic stress levels, habitat fragmentation, climate change (Woodward et al., 2010; Markovic et al., 2017) and the tremendous efforts for improvements (Knowler et al., 2004; Bernhardt et al., 2005), it is of fundamental importance to understand what environmental factors at what spatial scales (grain-size) are suitable predictors of which freshwater species' distributions.

The grain-size change is based on the most widely used hierarchical catchment ordering, the Strahler system (Strahler, 1964). Strahler order reflects the hierarchical level of each reach in the whole river network, with 1st order assigned to all reaches with no tributaries, 2nd order to the confluences of two first-order reaches, and so on. We focus on the “reach scale” (here, the catchment area of the river reach where an occurrence of the freshwater fish was observed) and the corresponding 2nd to 5th order catchments (following the catchment nomenclature of the CCM2 pan-European catchments database; CCM version 2.1, de Jager and Vogt, 2010; for further details see <http://ccm.jrc.ec.europa.eu>, accessed on 10.11.2018). The study area includes the Danube River Basin. The aims of the present study are: (1) to test whether there is a decrease in the performance of SDMs with coarsening grain-size (i.e., catchment order) and (2) to quantify what environmental factors at

what catchment orders are important predictors of which fish species' distributions.

2. Materials and methods

2.1. Study area

The Danube is Europe's second largest river basin (801,463 km²) and is shared among more than 80 million people from 19 countries. It extends from Central Europe through the Balkans and drains to the Black Sea (Fig. A.1). The Danube River delta is one of the world's largest wetlands, rich in rare fauna and flora and inscribed on UNESCO's World Heritage List in 1991. Due to its large spatial extent and diverse relief, the Danube River Basin also shows great differences in climate. The summed annual precipitation ranges from more than 2300 mm in the high mountains to less than 400 mm in the delta region, while the mean annual discharge reaches 6460 m³s⁻¹ at the Danube delta in Romania (ICPDR, 2009). The Danube River Basin hosts over 2000 plant species, 40 mammals species and approximately 100 fish species, and is subject to increasing pressure including pollution from agriculture, industry and municipalities (ICPDR, 2009).

2.2. Environmental data

Choice of environmental factors was based on recent studies of fish species distributions (Buisson et al., 2008; Lassalle et al., 2010; Markovic et al., 2012; Isaak et al., 2017). To provide information on the performance of freshwater SDMs at multiple scales, and in particular for the potential applications where fine resolution survey data are not available, all environmental data were gathered from publicly available databases. A total of 22 environmental variables emerged representing topographic, climatic and anthropogenic effects on ecosystems (Table 1). Topographic data were extracted from the CCM2 pan-European catchments database (CCM version 2.1, de Jager and Vogt, 2010). Climatic data were extracted from the Worldclim 30 arc-second (approx. 1 km × 1 km) gridded information (Hijmans et al., 2007). As a measure of anthropogenic pressure, we used land cover information extracted from the CORINE land cover database (EEA, 2011) and the number of inhabitants per area (Population) based on the Global Rural-Urban Mapping Project (GRUMP, version1) (available at <http://sedac.ciesin.columbia.edu/gpw>, accessed on 10.11.2018). Although different

aspects of the hydrologic regime, such as flow magnitude, frequency, timing and variability may impair ecological success of particular life stages of freshwater species (Domisch et al., 2015; Markovic et al., 2017), the availability of the discharge data is restricted even in the developed countries. The hydrologic predictors were not used as the aim of the study was to undertake performance assessment for SDMs that lacked fine grain survey data.

2.3. Hierarchical catchment orders

The analyses and modelling were conducted for five different catchment orders (WaterShed Order WSO1 to WSO5, Fig. A.1) based on the Strahler order of the river reaches from the CCM2 pan-European catchments database (CCM version 2.1, de Jager and Vogt, 2010). The 1st order catchments (WSO1) were defined according to the drainage areas of the individual river reaches, while higher order catchments (WSO2 – WSO5) result from groupings of the lower-order catchments in a hierarchical way. The “WSO1” to “WSO5” nomenclature originated from the CCM2 pan-European catchments database and was kept here for consistency. Catchment order increase was reflected in an increasing catchment size, with an average ranging from 12 km² for the WSO1 to 2148 km² for the WSO5. For each WSO, environmental factors were calculated by averaging gridded data across the corresponding catchment areas (Fig. A.1).

2.4. Fish data

Species occurrence data for the Danube River Basin were provided by Biofresh (www.freshwaterplatform.eu, Schinegger et al., 2016) for 1364 sites (Fig. A.1). Fisheries data were sampled by either single-pass or double-pass electrofishing between 1985 and 2002. To ensure an accurate estimate of the species distributions, only species with a minimum of 50 occurrences at the largest analysed scale, the WSO5, were included into the analysis (cf. Coudon and Gégout, 2007). Eight fish species were included in the study and ecological characterization followed Kottelat and Freyhof (2007): the bleak (*Alburnus alburnus*) is a small cyprinid, and prefers open waters of lakes and medium to large rivers. The stone loach (*Barbatula barbatula*) is usually found in medium-sized rivers with gravel to stone bottom. *Barbus barbus* is a fish of the cyprinid family preferably inhabiting faster flowing, summer-warm, medium to large-sized rivers. Gudgeons of the genus *Gobio*, here represented by *Gobio obtusirostris*, are riverine cyprinids, too, which tolerate – in contrast to barbels – lower flow velocities and finer spawning substrates. The bullhead (*Cottus gobio*) inhabits cold, clear and fast-flowing water of small streams to medium-sized rivers. The roach (*Rutilus rutilus*) is a small fish of the cyprinid family mainly found in nutrient-rich large to medium sized lowland rivers and backwaters. The trout (*Salmo trutta*) is a species of salmonid fish preferring cold, well-oxygenated streams in the mountainous areas. The chub (*Squalius cephalus*) is a fish of the cyprinid family found in slow-flowing lowland rivers, very small mountain streams, and in large streams of barbel zone.

For each of the five WSOs, fish occurrence data were aggregated to presence/absence information (hereafter called “catchment-scale mapping”). The final number of catchments of each particular order and their spatial arrangement was thus directly constrained by fish data availability. With the WSO increase, the number of catchments under consideration decreased (from 1363 for the WSO1 to 126 for the WSO5), with a corresponding increase in species prevalence (Table 2). We note here that, because of the dendritic structure of river networks, catchment-scale mapping is more appropriate for freshwater species than the point-to-grid mapping, used for mapping terrestrial species’ occurrences (see Fagan, 2002). In addition, given that catchments serve as units for freshwater management and conservation (commonly referred to as the Catchment Based Approach – CaBA, see DEFRA, 2013), catchment-scale mapping of freshwater species’ occurrences ensures

Table 2

Freshwater fish species and their prevalence. The total number of catchments for the studied scales (WSO1-WSO5) is indicated in parentheses.

Name	Code	WSO1	WSO2	WSO3	WSO4	WSO5
		(1363)	(1059)	(681)	(350)	(126)
<i>Alburnus alburnus</i>	Albual	0.20	0.22	0.29	0.38	0.62
<i>Barbatula barbatula</i>	Barbbr	0.31	0.33	0.34	0.35	0.45
<i>Barbus barbus</i>	Barbba	0.20	0.21	0.23	0.28	0.45
<i>Cottus gobio</i>	Cottgo	0.44	0.42	0.43	0.48	0.51
<i>Gobio obtusirostris</i>	Gobris	0.31	0.33	0.36	0.37	0.51
<i>Rutilus rutilus</i>	Rutiru	0.22	0.25	0.30	0.38	0.58
<i>Salmo trutta</i>	Saltta	0.71	0.70	0.72	0.74	0.69
<i>Squalius cephalus</i>	Squace	0.44	0.46	0.49	0.53	0.72

Ecological characterization followed Kottelat and Freyhof (2007): *A. alburnus* prefers medium to large rivers; *B. barbatula* prefers medium-sized rivers; *B. barbus* prefers faster flowing medium to large-sized rivers; *C. gobio* inhabits cold and fast-flowing water of small streams to medium-sized rivers; *G. obtusirostris* tolerate lower flow velocities; *R. rutilus* prefers nutrient-rich large to medium sized lowland rivers and backwaters; *S. trutta* prefers cold, well-oxygenated streams in the mountainous areas; and *S. cephalus* prefers slow-flowing rivers and very small mountain streams.

compatibility between the management and the analysis scales (Lévêque et al., 2008; Markovic et al., 2017).

2.5. Data analysis and modelling

Univariate strength of the environmental predictor variables was quantified using the weight of evidence (WOE) and information value (IV) concepts as implemented in the R (R Development Core Team, 2018) library “Information” (Larsen, 2016). While WOE describes the relationship between a predictor variable X (environmental predictors listed in Table 1) and a binary dependent variable Y (here species occurrence in a particular catchment), information value measures the strength of the $Y - X$ relationship. Specifically, if b_i , $i = 1, \dots, k$, denote k discrete bins for the predictor X , then, the strength of the predictor in describing Y can be quantified as $\sum_i (P(X \in b_i | Y = 1) - P(X \in b_i | Y = 0)) \times WOE_i$ (cf. Larsen, 2016). As such, information value is suitable for the initial predictor selection, i.e., for comparing the predictive power of the environmental factors. The information value was used to reduce the parameter number used in the modelling and thus avoid possible overfitting. Specifically, among several strongly correlated parameters (with pairwise correlations above |0.75|), the parameter with the largest information values was used in the modelling.

Species distribution modelling was performed using Generalized Additive Models (GAM) (R library “gam”; Hastie, 2005). GAM is a non-parametric extension of generalized linear methods, and is widely used for modelling current and future distribution patterns of fish species. Previous investigations using various SDMs have shown that GAM-, GLM- (Generalized Linear Models) and regression tree based SDMs have similar validation performance. The first two kinds of models had also similar calibration performance, while regression tree based SDMs tended to overfit during the calibration phase (Markovic et al., 2012). The improved performance of consensus or ensemble methods in providing more accurate and robust projections of species distribution have been already demonstrated (Marmion et al., 2009; Buisson et al., 2010; Lauzeral et al., 2013). However, the main objective of this study was determining the effect and importance of variation in spatial scale rather than the performance of different SDMs. Moreover, GAMs are very flexible models, and in contrast to the majority of SDMs, have well performance at high collinearity (Dormann et al., 2013). Therefore, we focussed only on GAM based SDMs, but acknowledge the importance of using multiple SDMs when the study goal is predicting future species distribution patterns (see Markovic et al., 2012; Meller et al., 2014).

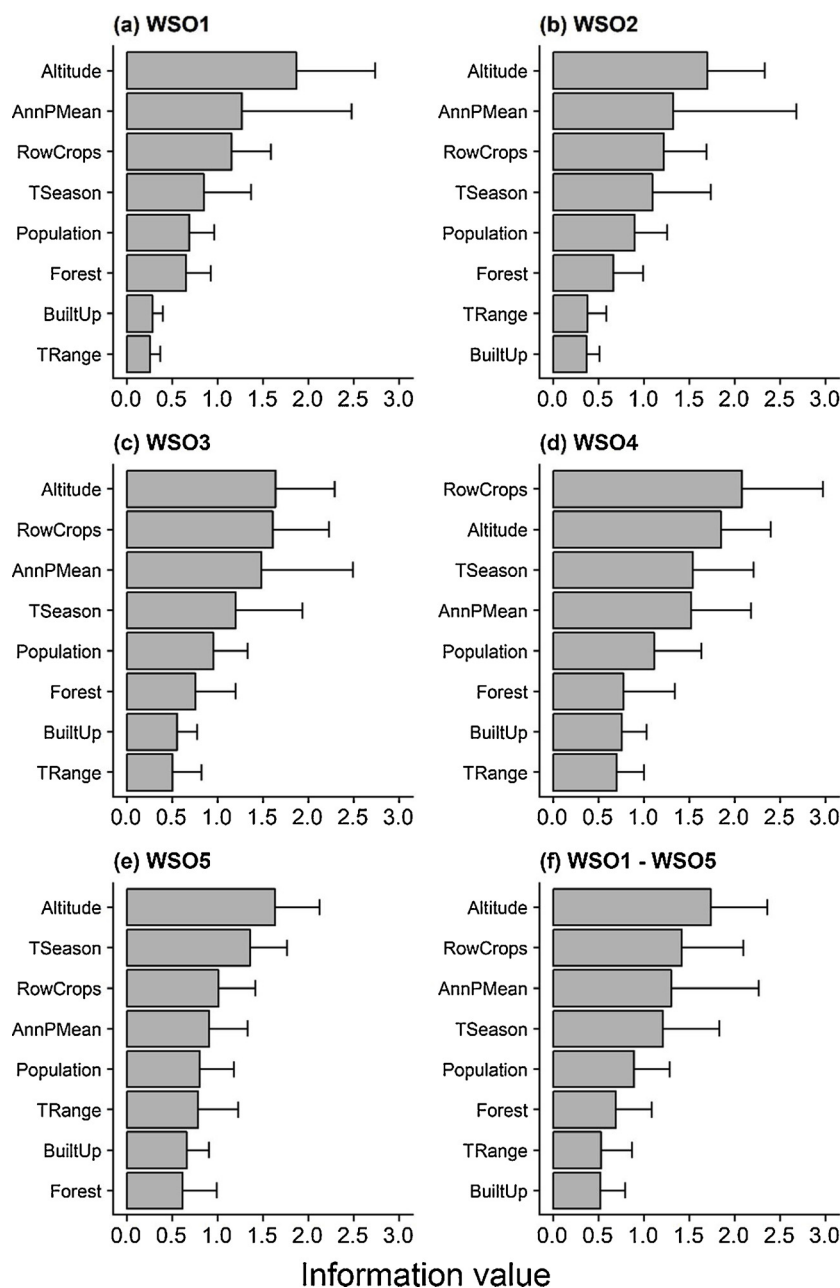


Fig. 1. Information value as a measure of the univariate strength of the environmental predictor variables. For convenience, the environmental predictors are ordered according to their information value. (a)–(e) For each of the five spatial scales the information value was calculated as an average predictive strength of the environmental predictor across all studied species; (f) the overall information value of an environmental predictor was calculated as an average predictive strength across all scales and all considered species. Error bars represent one standard deviation of the estimates.

Species occurrence probabilities resulting from GAM application were transformed to presence/absence information using the thresholds, which maximize both sensitivity (the true positive rate) and the specificity (the true negative rate). We applied random splitting of the fish data ten times into calibration (70%) and validation (30%), i.e., each of the ten models was calibrated using a different 70% data sample and validated using the remaining 30%. The repetitive modelling procedure allowed for quantifying the uncertainty of the estimates. Agreement between the observed and modelled species distribution patterns was quantified by sensitivity and specificity, while the performance of the calibrated models was estimated using the Area under the Receiver Operator Curve (AUC) and the True Skill Statistic (TSS) (Allouche et al., 2006). The use of the different statistical measures (here, sensitivity, specificity, AUC and TSS) was necessary to ensure

that sensitivities of the individual measures to prevalence and scale (see Lobo et al., 2008) were not misleading. An AUC of 0.5 and a TSS of 0 indicate that a model has no discriminatory power, while an AUC or TSS of 1 indicate that presences and absences are perfectly discriminated. The search for a parsimonious model involved analyses of the model improvement based on the Akaike Information Criterion (AIC) through simultaneous forward and backward predictor selection.

To quantify the relative predictor importance, the variance partitioning method by Lindeman et al. (1980), implemented within the R library “relaimpo” (Grömping, 2006) was used. The advantage of the variance partitioning method by Lindeman et al. (1980) is that it considers sequential sums of squares over all predictor permutations, and thus considers the inter-correlation effects among the individual predictors, with high predictor relative importance not necessarily

implying causation.

3. Results

3.1. Information value and correlations

The mean catchment elevation (Altitude) manifested the highest univariate strength in describing distributions of the studied fish species across the studied spatial scales (Table A.1 and Fig. A.2f). Specifically, for WSO1, WSO2 and WSO5, Altitude had by far the highest information value (IV), and at WSO4 the second highest. At WSO3, eight predictors (including Altitude) had almost equally high IV. Consequently, Altitude had to be retained as the predictor at all scales, while predictors with either marginal IV (Length, PSeason, Grassland and Pastures) or extremely high correlations (-0.99 to -0.93) with Altitude (Slope, Gradient, AnnTMean, AnnTMax, TAprSep, TMaiAug and WettTMean) were omitted from the analyses for parsimony reasons (Table A.1 and Fig. A.2). It is noteworthy that the pairwise correlation between Altitude and several variables describing temperature (AnnTMean, AnnTMax, TAprSep, TMaiAug and WettTMean) increased with scale. The remaining temperature variables TRange (mean IV = 0.52) and TSeason (mean IV = 1.21) (Table A.1), however, were included in order to account for direct temperature effects and not solely indirect given by the often-used temperature surrogate Altitude, whereby correlations between Altitude and TRange as well as TSeason were still high for catchment scales above WSO2 (-0.90 to -0.74) (e.g. Fig. A.3). We note that GAMs are very flexible models and still have well performance at high collinearity. For WSO1 and WSO2, however, pairwise correlations of these variables with Altitude were below $|0.75|$ (-0.74 to -0.54). Mean information values of further climatic variables, i.e. PMaiAug, PAprSep and AnnPMean, were close to each other (1.30–1.42) (Table A.1). High pairwise correlations among the three precipitation parameters and a lower correlation for WSO1 – WSO5 between AnnPMean and each of the already chosen variables, especially Altitude, led to the exclusion of PAprSep and PMaiAug. A low mean information value led also to the exclusion of the last precipitation variable PSeason (0.43) (Table A.1).

Among the anthropogenic variables, the area under row crops (RowCrops) emerged as the most valuable in describing distributions of the studied fish species. The high correlation between RowCrops and the second most valuable anthropogenic variable Agriculture (0.78 – 0.95) (e.g. Fig. A.3) has led to the omission of Agriculture. While the remaining anthropogenic variables BuiltUp, Forest and Population were included in the final variable set because of low pairwise correlations among the included variables and high information values within the anthropogenic category, the variables Grassland and Pastures were excluded due to low information values (Table A.1).

The final variable set consequently consisted of the topographic variable Altitude, the climatic variables TRange, TSeason and AnnPMean, and the anthropogenic variables RowCrops, Forest, BuiltUp, and Population (Fig. 1 and Table A.2).

3.2. Model performance

Using the selected eight predictor variables (Table A.2), we pursued two distinct model fitting approaches across all catchment scales and species: Modelling species distributions with GAM by (a) keeping the predictor number ($n = 8$) constant across the scales, and (b) using simultaneous forward and backward predictor selection in order to obtain a parsimonious model. When keeping all predictors across all scales, the overall validation performance was highly accurate with a mean AUC of 0.88 and a mean TSS of 0.63 across all catchment scales and species (Table A.3). Similar results held for the models with simultaneous forward and backward predictor selection where a mean validation AUC of 0.88 and a mean validation TSS of 0.62 (Table 3) were obtained. The highest performance values for both model fitting

Table 3

Mean and standard deviation (sd) for validation AUCs and TSSs of the multi-variate SDMs with simultaneous forward and backward predictor selection across all studied scales (WSO1–WSO5). Performance was assessed by using repeated random splitting (ten times) of the fish data into calibration (70%) and validation (30%).

	Accuracy measure		WSO1	WSO2	WSO3	WSO4	WSO5
<i>A. alburnus</i>	AUC	mean	0.90	0.90	0.91	0.91	0.91
		sd	0.01	0.02	0.02	0.03	0.05
	TSS	mean	0.61	0.65	0.68	0.71	0.69
		sd	0.04	0.04	0.04	0.08	0.10
<i>B. barbatula</i>	AUC	mean	0.83	0.86	0.87	0.87	0.77
		sd	0.02	0.02	0.01	0.02	0.05
	TSS	mean	0.53	0.58	0.60	0.59	0.47
		sd	0.05	0.04	0.04	0.03	0.12
<i>B. barbus</i>	AUC	mean	0.86	0.85	0.84	0.83	0.76
		sd	0.02	0.03	0.03	0.04	0.07
	TSS	mean	0.57	0.51	0.53	0.46	0.37
		sd	0.07	0.05	0.06	0.06	0.15
<i>C. gobio</i>	AUC	mean	0.83	0.86	0.86	0.90	0.85
		sd	0.01	0.01	0.02	0.02	0.05
	TSS	mean	0.53	0.57	0.56	0.67	0.57
		sd	0.02	0.03	0.04	0.05	0.13
<i>G. obtusirostris</i>	AUC	mean	0.89	0.89	0.89	0.89	0.81
		sd	0.02	0.01	0.02	0.02	0.08
	TSS	mean	0.64	0.64	0.60	0.63	0.46
		sd	0.03	0.05	0.06	0.07	0.19
<i>R. rutilus</i>	AUC	mean	0.90	0.91	0.90	0.91	0.85
		sd	0.01	0.02	0.02	0.02	0.06
	TSS	mean	0.66	0.69	0.69	0.71	0.64
		sd	0.03	0.05	0.04	0.05	0.15
<i>S. trutta</i>	AUC	mean	0.94	0.95	0.97	0.98	0.96
		sd	0.01	0.01	0.01	0.01	0.04
	TSS	mean	0.75	0.76	0.82	0.89	0.82
		sd	0.03	0.02	0.04	0.04	0.10
<i>S. cephalus</i>	AUC	mean	0.87	0.87	0.87	0.87	0.83
		sd	0.01	0.02	0.02	0.03	0.07
	TSS	mean	0.60	0.59	0.56	0.57	0.49
		sd	0.03	0.04	0.05	0.08	0.15

approaches were observed for *S. trutta* at WSO4 (AUC = 0.98) (Table 3 and Table A.3). Overall, when coarsening the grain size (i.e. increasing catchment order), both a slight model degradation and a slight model improvement was observed (Figs. A.4 and A.5). In addition, the increasing catchment order was accompanied by an uncertainty increase manifested by an increase in the standard deviation of the validation sensitivity, specificity, AUC and TSS (Figs. 2 and A.6).

3.3. Predictor importance per species and catchment order

Predictor importance patterns varied across species and catchment orders (Figs. 3 and 4, Table A.4–A.5, Fig. A.7–A.14). High variation throughout species is shown in Fig. 3 by high standard deviations for all variables. For both model fitting approaches and all species, highest predictor importance was distributed among the variables Altitude (20.1%–37%), TSeason (13.5%–22.5%), AnnPMean (14.1%–27.5%) and RowCrops (4.3%–25.3%) (Fig. 3). In general, lowest predictor importance was present for TRange, Forest, BuiltUp and Population. However, TRange was shown to have an increasing overall mean importance with increasing catchment order (5.1%–13.4%) and Forest an overall nearly constant importance across all scales at around 10%, both inferred from the first model fitting approach, which kept the predictor number constant. Using simultaneous forward and backward predictor selection implied a mean predictor importance below 9% for the four least important variables across all scales (Fig. 3). Trends in the predictor importance for the four most important variables across the five catchment orders varied depending on whether the predictors were kept constant across scales or automatically selected. The first approach identified a downward trend of the importance for Altitude with

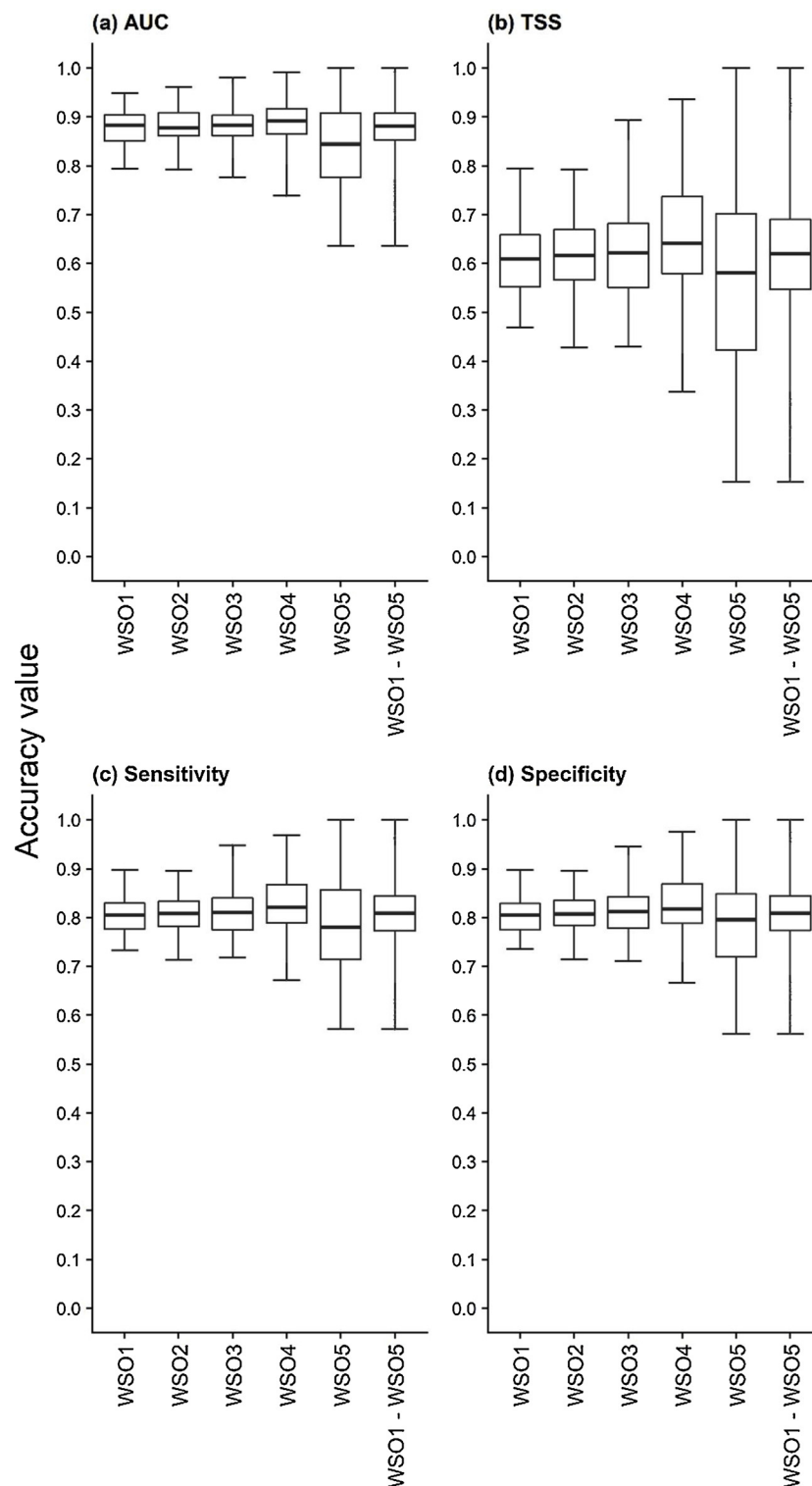


Fig. 2. Mean performance measures of the multivariate SDMs with simultaneous forward and backward predictor selection across the five studied scales (WSO1 – WSO5). Performance was assessed by using repeated random splitting (ten times) of the fish data into calibration (70%) and validation (30%) and calculating the (a) AUC (area under the receiver operating curve), (b) TSS (true skill statistic), (c) sensitivity and (d) specificity.

increasing scale, whereas the second approach delineated an increasing importance with increasing scale. Trends of predictor importance along the studied scales inferred from both approaches were similar for TSeason, which could be described by an upward trend, whereas the importance for WSO4 and WSO5 deduced from the second approach was higher. Both model fitting approaches assigned high importance to AnnPMean from WSO1 to WSO4 and lower importance at WSO5. The

second approach, however, showed greater variation for AnnPMean. The predictor importance of RowCrops for both approaches revealed highest explanation power at local scales and lower importance at WSO4 and WSO5 (Fig. 3). Overall, at least one variable of the three considered categories (“topographic”, “climatic”, “anthropogenic”), respectively, was identified as important in describing fish species distributions.

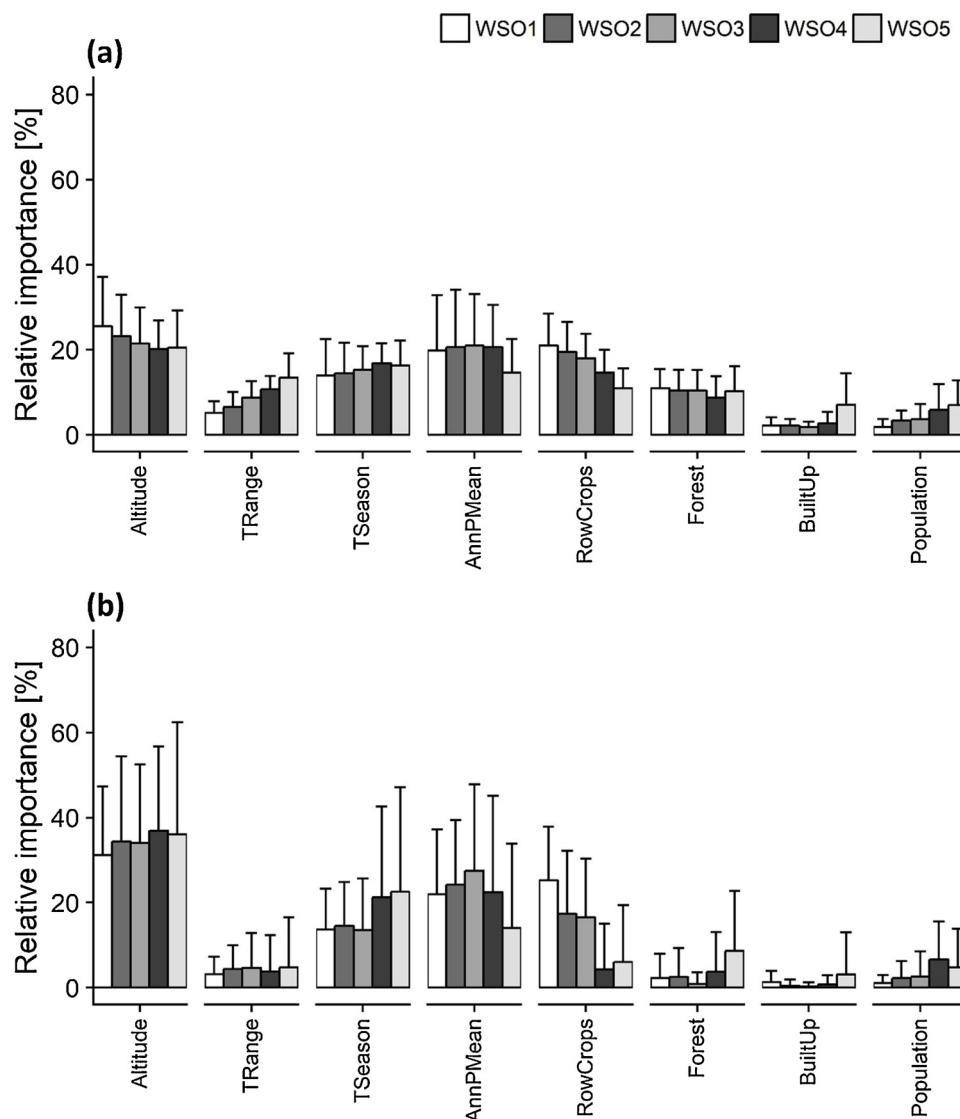


Fig. 3. Mean relative predictor importance resulting from the SDMs across all studied species and all studied scales (WSO1-WSO5) for two distinct model fitting approaches: (a) with keeping the predictor number constant across the scales, and (b) with simultaneous forward and backward predictor selection. Error bars represent one standard deviation of the estimates. We note that the figure represents the mean relative predictor importance based on all model runs per species and catchment order.

On the single species level, most important variables consisted in general of a subset of the overall identified four most important variables (Altitude, TSeason, AnnPMean, RowCrops) (Fig. 4, Tables A.4 and A.5, Fig. A.7–A.14). Few exceptions were for example *B. barbus* or *G. obtusirostris* where higher predictor importance with increasing scale of TRange for the first model fitting approach was observable (Figs. A.9 and A.11). Moreover, some of the four most important predictors were only selected at certain catchment orders. As such, TSeason as explanatory variable for the distribution of *B. barbus* had a predictor importance above 10% only for WSO4 and WSO5, whereas for WSO2 and WSO3 the variable was not selected as predictor (Fig. 4a and Table A.5). However, Altitude was dominating the predictor importance pattern of *B. barbus* from WSO1 to WSO5 and the distribution of *B. barbus* at the local levels WSO1 to WSO3 was strongly influenced by RowCrops (Fig. 4a and Table A.5). This pattern was also identified by the first modelling approach (Fig. A.9a and Table A.4). In addition, the anthropogenic variable Population, here generally identified as variable with low importance, had an importance of 26.8% at WSO4 for models with automatic predictor selection. Such occasional high importance of variables with generally low importance could also be observed for

other considered species. For *G. obtusirostris*, Altitude, TSeason, AnnPMean and RowCrops emerged as the key factors (Figs. 4b and A.11, and Tables S4 and S5). Similarly, for *S. trutta*, Altitude, TSeason, AnnPMean and RowCrops were most often selected as predictors across all scales (Figs. 4c and A.13, and Tables S4 and S5).

4. Discussion

The majority of models of fish distribution patterns across the studied spatial scales were highly accurate for both used model fitting approaches. The negligible effect of the catchment order on the mean performance of freshwater SDMs is in line with the results of Guisan et al. (2007) who have shown that coarsening study grain-size through grid-cell size increase, does not have substantial effects on the performance of terrestrial SDMs. However, our study additionally shows that the variance of the validation performance measures increases with increasing scale. The high accuracy of the SDMs across all studied catchment orders confirms the overall appropriateness of the considered environmental factors, and is also a reflection of the predictive ability of the applied statistical methodology, i.e. GAM. As such, we

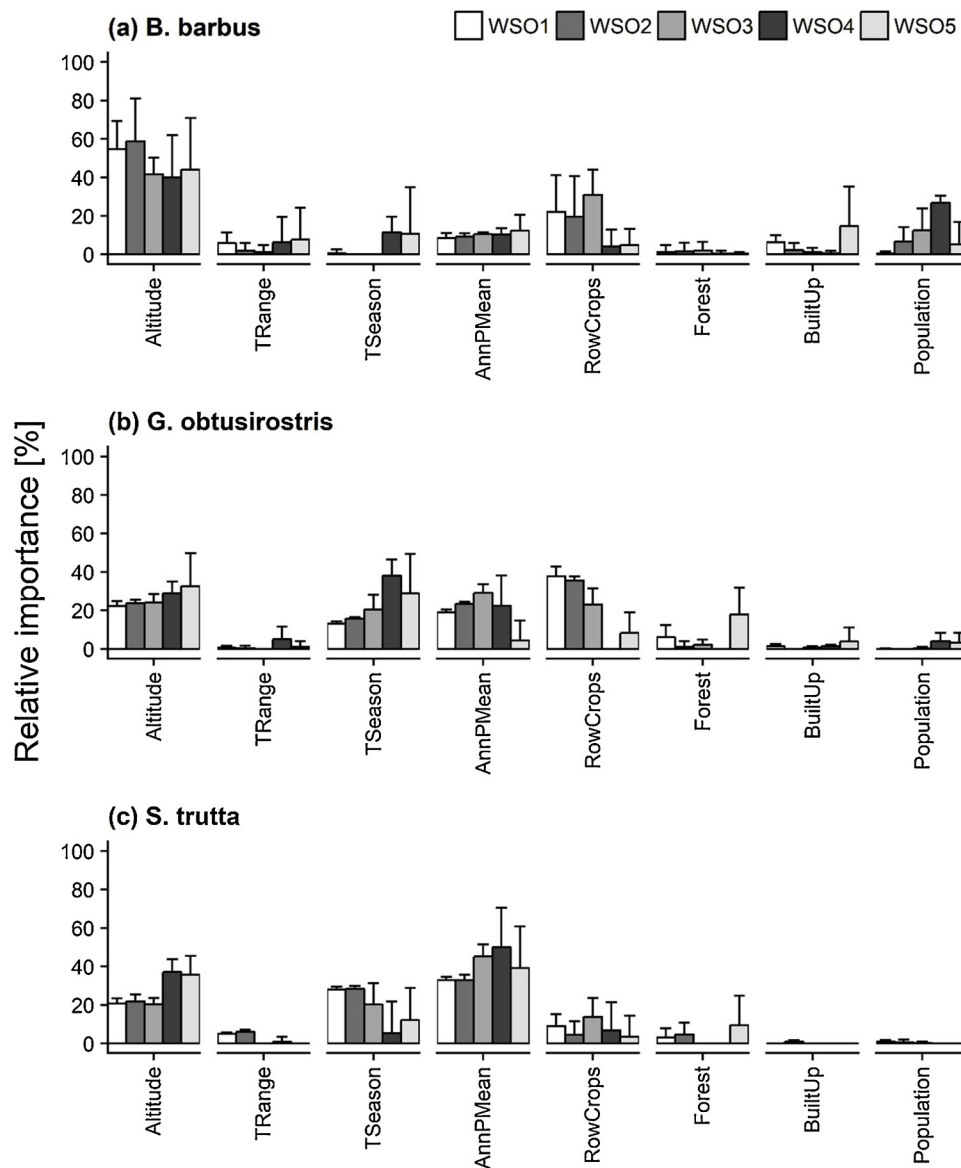


Fig. 4. Relative predictor importance in describing distribution patterns of (a) *B. barbus*, (b) *G. obtusirostris* and (c) *S. trutta* across the studied scales (WSO1-WSO5) inferred from the multivariate SDMs with simultaneous forward and backward predictor selection. Error bars represent one standard deviation of the estimates. We note that the figure represents the mean relative predictor importance based on all model runs per catchment order.

note that adding new parameters to our models would rather result in an overfitting than in a meaningful improvement in the model accuracy.

When looking at the predictive ability of each individual environmental factor, the topographic variable Altitude, the climatic variables TSeason and AnnPMean, and the anthropogenic variable RowCrops tend to be the most important predictors irrespective whether or not an automatic predictor selection was used. Altitude, which can be seen as a surrogate for temperature, and climatic factors are well known to influence fish species distributions (e.g. Kuemmerlen et al., 2014). The identification of RowCrops as important predictor aligns with previous studies stressing the importance of surrounding landscape on the in-stream ecosystem structure and function (Fausch et al., 2002; Linke et al., 2008). The link between the area covered by RowCrops and fish distribution patterns was argued by Strayer et al. (2003) as high nitrate flux leading to a high level of aquatic plant cover which in turn leads to low fish species richness. Similarly, intensely used crop lands with its bare soils free of weeds suffer from surface erosion and thus, provide a continuous source of fine sediments leading to clogging and siltation of

coarse substrates followed by the decline of gravel-spawning riverine fish species (Soulsby et al., 2001; Lapointe et al., 2004; Greig et al., 2005; Jensen et al., 2009). Overall, our results, especially the differences in predictor importance trends across scales of the two approaches and the corresponding variable selection, suggest *species dependent selection* of factors describing species distributions, paired with appropriate monitoring of effects and management of mitigation activities tailored to ensure species' long-term persistence.

The relative importance of the factors used to describe species distribution patterns is known to vary across spatial scales (Jackson et al., 2001; Blackburn and Gaston, 2002; Tudesque et al., 2014). At the same time, the relative importance of the spatial scale and the environmental factors used in species distribution modelling is largely determined by a combination of species range and species prevalence (Hopkins and Burr, 2009). Our results indicate that for *B. barbus* the relative importance of anthropogenic pressure – manifested by RowCrops, Population and BuiltUp – is similar to or higher than the importance of the considered three climatic factors. *B. barbus* is known to be especially sensitive to damming, river regulation and fine sediment input

indicating that the results of the SDMs well align with the ecological classification of this species (Kottelat and Freyhof, 2007). Also, *B. barbus* seems more susceptible to the effects of land-use on water quality and spawning habitat than *S. trutta* (which has similar environmental requirements). However, this seeming relation simply results from higher cumulative anthropogenic pressure at the lower elevated river reaches preferably colonised by barbel. This is further confirmed by the finding that the influence of RowCrops at local scales better describes the distribution of gudgeons of the genus *Gobio*, which prefer cold and clean water. Our finding of the dominant role of climatic variables in shaping distribution patterns of the fish species studied for all catchment orders, contradicts the suggestions that climate related factors may be good predictors of species distributions only at the macro-scale (e.g. Pearson and Dawson, 2003). However, for robust conclusions to be made further integrated climate impact modelling would be needed.

Although it contributes significantly to the understanding of freshwater fish patterns at various scales, our modelling framework has a number of limitations. One limiting aspect of our study framework lies in the fact that the selected set of catchment orders and the predictor dataset used do not account for the effects of spatial arrangements of local habitats and spatial autocorrelation effects, as well as habitat peculiarities with respect to species requirements at different life stages. Specifically, proximal environment experienced by a species is not necessarily reflected in the geographic scales used to calculate the environmental factors. As such, the probability of species' occurrence resulting from the calibrated models provides the information on the potential habitat suitability of catchments at different hierarchical levels. This suitability does not consider the accessibility of habitats required for particular life stages, or temporal resource availability and adequateness.

Another weakness is the use of multi-decadal averages of climatic factors in the model calibration process. Climate information for the last 30–50 years may not adequately reflect species environmental tolerance ranges. Moreover, the tolerance limits are rarely rigidly fixed, as they might depend on lifetime experience, the developmental stage of individuals within species and the combined effect of various abiotic and biotic factors. Consequently, in order to account for temporal and spatial aspects of species distribution patterns, researchers will need to disentangle species information according to life stages, the life strategy and temporal mismatch of predictors to response, as well as environmental information according to temporal variability scales of the physical processes governing individual environmental properties. Time, space and hierarchical level are the fundamental axes of scale, and since many physical and ecological phenomena are related in space versus time (Wu and Li, 2006) temporal information should be accounted for in models of species distributions (see also Soranno et al., 2014).

5. Conclusion

Coarsening study grain-size through catchment order increase are shown to be of minor influence on the mean performance of freshwater fish SDMs. However, the uncertainty of the estimates increases with scale. We highlight the importance of using anthropogenic effects as population and land cover related predictors when addressing species sensitive to pollution such as *B. barbus*. Specifically, our results indicate that the model complexity and the importance assigned to environmental drivers of fish distributions are both catchment order dependent and species dependent. The latter emphasises importance of compatibility between the scale of factor importance and species conservation management scale.

With the current threat to freshwater biodiversity and lack of information on which to base freshwater conservation strategies in many regions of the world, in particular less-developed countries, our results have useful implications for predicting distributions of species and

conservation planning in areas where fine grain survey data on species occurrence are unavailable. In particular, our results indicate that the use of broader scale species and environmental data (i.e. up to WSO5 catchment scale data), does not significantly affect the performance of SDMs. We remind here that the ecological scales of relevance of species and environmental variables should be matched (Hurlbert and Jetz, 2007), although the increased availability of high resolution environmental data is tempting towards studying species distributions at finer resolutions than the species survey data, to avoid misleading conservation assessments. In addition, in view of the extreme global pressure of climate change on freshwater ecosystems and the importance of the climatic factors on distribution patterns of the studied freshwater fish species even at the smallest studied scale (reach scale), our findings underline the necessity of efforts to continuously re-assess the potential effects of climate change on suitability of catchments for freshwater fish species. Given the high vulnerability of freshwater ecosystems paired with limited dispersal ability of strictly aquatic species such as fish, addressing environmental effects across multiple (spatial and temporal) scales is central to effective species conservation and timely identification of potentially detrimental changes in habitat suitability.

Acknowledgements

This work was supported by the DFG under Grant MA 6593/2-1 and by the Horizon 2020 ECOPOTENTIAL project (ref. 641762).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ecolmodel.2019.05.006>.

References

- Allan, J.D., et al., 1997. The influence of catchment land use on stream integrity across multiple spatial scales. *Freshw. Rev.* 37, 149–161.
- Allouche, O., et al., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232.
- Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688.
- Bernhardt, E.S., et al., 2005. Synthesizing U.S. River restoration efforts. *Science* 308, 636–637.
- Blackburn, T.M., Gaston, K.J., 2002. Scale in macroecology. *Global Ecol. Biogeogr.* 11, 185–189.
- Brotons, L., et al., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27, 437–448.
- Buisson, L., et al., 2008. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecol. Freshw. Fish* 17, 244–257.
- Buisson, L., et al., 2010. Uncertainty in ensemble forecasting of species distribution. *Glob. Change Biol. Bioenergy* 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>.
- Coudon, C., Gégout, J.C., 2007. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. *J. Veg. Sci.* 18, 517–524.
- de Jager, A.L., Vogt, J.V., 2010. Development and demonstration of a structured hydrological feature coding system for Europe. *Hydrol. Sci. J. Des Sci. Hydrol.* 55, 661–675.
- DEFRA, 2013. Catchment Based Approach: Improving the Quality of Our Water Environment. Policy paper available at: Department for Environment, Food & Rural Affairs. <https://www.gov.uk/government/publications/catchment-based-approach-improving-the-quality-of-our-water-environment>.
- Domisch, S., et al., 2013. Choice of study area and predictors affect habitat suitability projections, but not the performance of species distribution models of stream biota. *Ecol. Model.* 257, 1–10.
- Domisch, S., et al., 2015. Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundam. Appl. Limnol.* 186, 45–61.
- Dormann, C.F., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- EEA, CORINE Land Cover, 2011. Version 13, Copenhagen: European Environmental Agency (EEA). Available at: www.eea.europa.eu/publications/COROLandcover.
- Efron, B., Hastie, T., 2016. *Computer Age Statistical Inference*. Cambridge University Press, Cambridge, UK.
- Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. *Ecography* 32,

- 66–77.
- Esselman, P.C., Allan, J.D., 2010. Relative influences of catchment- and reach-scale abiotic factors on freshwater fish communities in rivers of northeastern Mesoamerica. *Ecol. Freshw. Fish* 19, 439–454.
- Fagan, W.F., 2002. Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology* 83, 3243–3249.
- Fausch, K.D., et al., 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52, 483–498.
- Franklin, J., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, New York, NY.
- Greig, S.M., et al., 2005. Fine sediment accumulation in salmon spawning gravels and the survival of incubating salmon progeny: implications for spawning habitat management. *Sci. Total Environ.* 344, 241–258.
- Grenouillet, G., et al., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34, 9–17.
- Grömping, U., 2006. Relative importance for linear regression in R: the package relaimpo. *J. Stat. Softw.* 17, 1–27.
- Guisan, A., et al., 2007. Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.* 13, 332–340.
- Hastie, T.J., 2005. *Generalized Additive Models*. R Package Version 0.94.
- Hijmans, R.J., et al., 2007. *WorldClim Version 1.4*. Museum of Vertebrate Zoology. Available at: University of California. <http://www.worldclim.org/>.
- Hopkins, R.L., Burr, B.M., 2009. Modeling freshwater fish distributions using multiscale landscape data: a case study of six narrow range endemics. *Ecol. Model.* 220, 2024–2034.
- Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13384–13389.
- ICPDR, 2009. *The Danube River Basin: Facts and Figures*. International Commission for the Protection of the Danube River (ICPDR), Vienna.
- Isaak, D.J., et al., 2017. Big biology meets microclimatology: defining thermal niches of ectotherms at landscape scales for conservation planning. *Ecol. Appl.* 27, 977–990.
- Jackson, D., et al., 2001. What controls who is where in freshwater fish communities: the roles of biotic, abiotic and spatial factors? *Can. J. Fish. Aquat. Sci.* 58, 157–170.
- Jensen, D.W., et al., 2009. Impact of fine sediment on egg-to-fry survival of Pacific salmon: a meta-analysis of published studies. *Rev. Fish. Sci. Aquac.* 17, 348–359.
- Knowler, D.J., et al., 2004. Valuing freshwater salmon habitat on the west coast of Canada. *J. Environ. Manage.* 69, 261–273.
- Kottelat, M., Freyhof, J., 2007. *Handbook of European freshwater fishes*. Cornol; Kottelat, Berlin: Freyhof 646.
- Kuemerlen, M., et al., 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. *Ecol. Model.* 277, 77–86.
- Lapointe, M.F., et al., 2004. Interactive effects of substrate sand and silt contents, redd-scale hydraulic gradients, and interstitial velocities on egg-to-emergence survival of Atlantic salmon (*Salmo salar*). *Can. J. Fish. Aquat. Sci.* 61, 2271–2277.
- Larsen, K., 2016. R Package 'Information'. Version 0.0.9.
- Lassalle, G., et al., 2010. Global warming impacts and conservation responses for the critically endangered European Atlantic sturgeon. *Biol. Conserv.* 143, 2441–2452.
- Lauzeral, C., et al., 2013. Spatial range shape drives the grain size effects in species distribution models. *Ecography* 36, 778–787.
- Lawson, C.R., et al., 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods Ecol. Evol.* 5, 54–64.
- Lévesque, C., et al., 2008. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia* 595, 545–567.
- Lindeman, R.H., et al., 1980. *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL, Scott, Foresman 444 p.
- Linke, S., et al., 2008. Irreplaceability of river networks: towards catchment-based conservation planning. *J. Appl. Ecol.* 45, 1486–1495.
- Lobo, J.M., et al., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151.
- Markovic, D., et al., 2012. Where are all the fish: potential of biogeographical maps to project current and future distribution patterns of freshwater species. *PLoS One* 7, e40530. <https://doi.org/10.1371/journal.pone.0040530>.
- Markovic, D., et al., 2017. Vulnerability of European freshwater catchments to climate change. *Glob. Change Biol. Bioenergy* 23, 3567–3580. <https://doi.org/10.1111/gcb.13657>.
- Marmion, M., et al., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15, 59–69.
- Meller, L., et al., 2014. Ensemble distribution models in conservation prioritization: from consensus predictions to consensus reserve networks. *Divers. Distrib.* 20, 309–321.
- Morán-Ordóñez, A., et al., 2016. Evaluating 318 continental-scale species distribution models over a 60 year prediction horizon: what factors influence the reliability of predictions? *Glob. Ecol. Biogeogr.* 26, 371–384.
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371.
- R Development Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Schneider, R., et al., 2016. *Metadata describing the European Fish Index Plus (EFI+) database*. *FMJ* 17, 1–12. <https://doi.org/10.15504/fmj.2016.17>.
- Soranno, P.A., et al., 2014. Cross-scale interactions: quantifying multi-scaled cause-effect relationships in macrosystems. *Front. Ecol. Environ.* 12, 65–73.
- Soulsby, C., et al., 2001. Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: a preliminary assessment. *Sci. Total Environ.* 265, 295–307.
- Strahler, A.N., 1964. Quantitative geomorphology of drainage basins and channel networks. In: Chow, V.T. (Ed.), *Handbook of Applied Hydrology*. McGraw-Hill, New York, pp. 4–74.
- Strayer, D.L., 2003. Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. *Ecosystems* 6, 407–423.
- Thuiller, W., 2004. Patterns and uncertainties of species range shifts under climate change. *Glob. Change Biol. Bioenergy* 10, 2020–2027.
- Tudesque, L., et al., 2014. Scale-dependent effects of land cover on water physico-chemistry and diatom-based metrics in a major river system, the Adour-Garonne basin (South Western France). *Sci. Total Environ.* 466–467, 47–55.
- Woodward, G., et al., 2010. Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philos. Trans. Biol. Sci.* 365, 2093–2106.
- Wu, J., Li, H., 2006. Concepts of scale and scaling. In: Wu, J., Jones, K.B., Li, H., Loucks, O.L. (Eds.), *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Dordrecht; Springer, pp. 3–16.