AMBER

www.amber.international

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 689682.

D2.1 Classification map of running waters considering fish community structure and barrier impacts

This is version 2.0 D2.1 `Classification map of running waters considering fish community structure and barrier impacts .This document is a deliverable of the AMBER project, which has received funding from the European Union's Horizon 2020 Programme for under Grant Agreement (GA) #689682.

History of changes

Version	Date	Changes	Pages
1.0	03 Jan 2018		
2.0	18 April 2018		

DISCLAIMER

Executive summary

This is version 2.0 of D2.1 `Classification map of running waters considering fish community structure and barrier impacts`. This document is a deliverable of the AMBER project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689682.

The core element of this deliverable is a GIS map that represents the Water Bodies classified according to the Fish Community Macro Habitat Types that support specific reference structure of the fish fauna. This is the first step towards establishing a benchmark for determining barrier impacts on fish communities. Such classification is necessary because we expect that barrier impacts will vary depending on the type of fish community.

Using available, Europe-wide databases containing information about environmental characteristics and electrofishing data we intended to analyse the relationship between the physical setting and expected, reference fish community structure. The European Intercalibration database gathered during the Intercalibration process conducted between 2006-2011 by European Commission Joint Research Centre - JRC (WFD Intercalibration 2011) was used for this purpose. The database contains over 5,000 fish records from 23 countries. Permissions to use the data have been received from 19 countries. To develop the European Fish Community Macro Habitat Types of riverine water bodies a the following abiotic data were used for statistical analysis: site coordinates, country, size of catchment, altitude, surficial geology, valley slope, bioregion and Strahler stream order. These macro scale attributes are insensitive to human induced alterations and therefore allow us to establish the physiographic relationships to biota representing reference state. The database includes also information about human disturbances that allowed us to select non-disturbed reference sites for model calibration.

The biological data used in the classification map include the estimated abundance of different fish species divided into Habitat use and Tolerance guilds (derived from studies conducted in an earlier EU-funded project, EFI+). The macrohabitat typology of each waterbody was then determined by the proportion composition of guilds found at reference sites, together with the environmental attributes of those sites. Two step cluster and discriminant analysis were used for statistical classification, resulting in an accurate model ($R=0.937$, $p<0.001$) of eight classes. An additional 9th class was added by expert assessment through splitting the last 8th class according to geographical location of the sites (Scandinavian vs. Mediterranean). For each of those Fish Community Macro Habitat Type (FCMacHT) classes a template structure model of Expected Fish Community representing Habitat Use and Tolerance (HUT) guilds proportions was developed using a Target Fish Community approach. The resulting pie-charts capture the biological differences between classes, and highlight the fact that for the majority of rivers the fish community is dominated by rheophylic (i.e. riverine) fish species.

A subsequent step Classification and Regression Tree (CART) model was used to explore the relationship between the macro scale environmental attributes and the FCMacHT class. This allowed us to assign data originating from disturbed sites. The model resulted in 92% correct classification.

The FCMacHT classes are plotted on the map for the waterbodies available in the European Databases. The base map used is a combination of the Water Information System for Europe Water Framework Directive reference spatial data sets (WISE WFD, EEA, 2017) and Catchment Characterisation and Modelling River and Catchment Database, version 2.1 (CCM2) (Vogt *et al.* 2007).

Author(s):

Piotr Parasiewicz (SSIFI), Paweł Prus (SSIFI), Kamila Belka (ERCE), Mikołaj Adamczyk (SSIFI), Małgorzata Łapińska (ERCE), Maciej Zalewski (ERCE), Katarzyna Suska (SSIFI), Janusz Ligięza (SSIFI), Jacek Szlakowski (SSIFI), Wouter van de Bund (JRC), Carlos Garcia de Leaniz (SU), Ian Dodkins (SU), Martyn Lucas (DU), Paul Kemp (SOTON), Kim Aarestrup (DTU), Simone Bizzi (POLIMI), Klemens Kaupert (IBK) for the AMBER consortium.

Acronyms

These acronyms are used throughout the text:

CART	Classification and Regression Tree (statistical classification method)
EFI+	new European Fish Index (method to assess ecological status of fish in European Rivers)
HUT	Habitat Use and Tolerance (fish guilds based around habitat, and based on EFI+)
FCMacHT	Fish Community Macro Habitat Type (physically defined habitats considered to be associated with fish)
IRS	International Reference Sites as defined in the Interalibration Database
NHC	Non-Hierarchical Cluster (a type of statistical analysis to create groups)
NDS	Non-Disturbed Sites (river sites considered to have little impact from human activity according to methodology applied in the Intercalibration Database)
PAM	Partitioning Around Medoids (a statistical method of defining classes within a model)
TFC	Target Fish Communities (expected fish species given no disturbance, for rehabilitation)
WFD	Water Framework Directive (European Freshwater Legislation)

Contents

1 INTRODUCTION

The aim of this deliverable is to establish a benchmark typology of riverine macrohabitats to measure the impact of barriers on fish communities. The algorithms developed in this deliverable predict the expected fish communities from macroscale physical characteristics of the water body.

Establishing biological reference conditions is a requirement of the Water Framework Directive, necessary for the determination of a site's ecological status. Most of the schemes currently available for defining reference conditions are structured by site-specific physical characteristics, and take into account the level of human induced alteration, assuming that sites with low pressures define reference sites, and that fauna sampled at such low pressure sites represent reference conditions for the site. Within the European Intercalibration exercise led by ECOSTAT, indices such as EFI+ (Extension of the European Fish Index) were calculated using this approach. Their reference values are based on sites with limited human pressure.

An alternative approach has been developed recently that aims to predict macrohabitat distribution supporting specific fish community structure in macroscale physiographic settings. The settings are described from the assemblage of physical attributes that are insensitive to anthropogenic pressures (e.g. watershed area, gradient or elevation). Deviation of local community from such expected assemblages can be easily measured and related to human induced alterations. A prime example of such an approach developed for Laurentian Great Lakes is presented by McKenna *et. al* 2006, Steen *et al.* 2008, Lyons *et al.* 2010, McKenna & Johnson 2011. It predicts the expected proportions of fish species and diversity index for Stream Reaches using neural networks and decision trees, and explained between 49 and 99% of variability in calibration data. Neighbouring Stream Reaches with predicted macrohabitats for similar communities are then aggregated into Fisheries Management and Conservation units, which are clustered into specific types. Since Stream Reaches are stream sections between two confluences with similar hydromorphological characteristics they are almost equivalent to Water Bodies as identified by the EU Water Framework Directive (2000) for Europe.

Fish surveys across large regions are affected by variation among surveyors, differences in methodology, seasonality, and year of survey. This affects the accuracy of fish community prediction models. To account for this, the predicted fish distributions can be fitted into generic distribution patterns, such as those proposed by biocomplexity models (Bak 1996). The Target Fish Community approach (Bain & Meixler 2007) uses the ranking of relative species abundance found in a survey to calculate expected proportions according to the biocomplexity model.

To further reduce regional variability among samples, it has been proposed to group species into habitat use guilds and estimate the quantity of a guild assemblage rather than a species. This has been used in Poland to determine environmental flows (Parasiewicz *et al.* 2015). It is an intermediate step towards developing a mesoscale habitat model for the Expected Fish Community. Such a model serves as a benchmark for the assessment of current and future spatio-temporal habitat distribution. In this deliverable of the AMBER project, we used an adaptation of the Polish procedure and apply it to the large data set that was previously gathered during the European intercalibration exercise.

2 MATERIAL & METHODS

2.1 Data sources

The core source of fishery data is a European Intercalibration database (IC) gathered during the intercalibration process conducted between 2006-2011 under the auspices of the European Commission Joint Research Centre – JRC (WFD Intercalibration 2011). This database contains information on 4561 fished sites in 22 countries. The database consists of several tables describing basic physical parameters for each site, characteristics of a given fishing campaign, and catch results. Data on anthropogenic pressures were also gathered, which permits to classify the sites into reference and disturbed sites.

The Intercalibration database was used with permission from JRC and data owners from 19 States involved in the process. A similar dataset for Poland, gathered during the auto-intercalibration process (Prus *et al.* 2016) was also used, with the permission of Chief Inspectorate for Environmental Protection, Poland.

From this data a subset was extracted consisting of 1099 sites classified as representing low anthropogenic pressure (non- disturbed) sites (NDS), and used for FCMacHT model calibration.
The metadata consists of selected physical variables for each site as well as the relative proportions of each species found in the sample. The selected physical variables of IC are as follows (**Table 1**):

- Geomorphic river type
- Size of catchment
- Altitude
- Geological typology
- Actual river slope
- Natural sediment
- Wetted width
- Floodplain connectivity (E_floodplain)
- Intercalibration Region (IC_Region)
- Stream Order (Strahler_SO)

However, the Intercalibration data set is limited to a sample of European rivers and data on the attributes considered above are not available with the same level of accuracy for all European waterbodies. Our goal, therefore, was to develop a predictive map of macrohabitat distribution that would be applicable across Europe. To supplement missing information we searched for an equivalent of the above physical variables in the public pan-European datasets (**Table 1**).

Table 1. Physical variables of the Intercalibration dataset (IC), their relative importance factor estimated by CART, and unified new physical variables used for CART analysis.

IC Variable	Description	Importance	New Variable Code	Description
Actual.river.slope	Slope of the river (‰) measured on a stretch of length corresponding to catchment size: 1 km (for catchment area up to 100 km²), 5 km (catchment area 100-1000 km²) and 10 km (catchment area above 1000 km²)	15	CCM2.W_SLOPE	Slope of the river reach derived from the CCM2 database, calculated following the formula: [CCM2.ALT_GRADIE]*10
Altitude	Altitude of a site (m above sea level)	14	CCM2.W_ALT	Mean altitude of the river segment as a mean altitude between the altitudes of its beginning and its end (m a.s.l.), calculated based on CCM2 following the formula: (Elevation of Fromnode + Elev of ToNode) / 2
Size.of.catchment	Size of the river catchment above the site (km²)	13	CCM2.W_CATCH	Size of catchment of the river segment calculated as a sum of the immediate catchment of the river segment and all catchments above the given river segments derived from the CCM2 database
IC_GROUP	Intercalibration regional group: Nordic, Lowland-Midland, Alpine-type Mountains, Mediterranean South-Atlantic, Danubian (WFD Intercalibration, 2011)	12	W_BIO_ATL W_BIO_BOR W_BIO_ALP W_BIO_CONT W_BIO_MED W_BIO_PANN W_BIO_STEP W_BIO_BLAC W_BIO_ANA W_BIO_ARC W_BIO_OUT (1-0 value fields)	Biogeographical regions of Europe: Atlantic, Boreal, Alpine, Continental, Mediterranean, Pannonian, Steppic, Black Sea, Anatolian, Arctic, outside Europe derived from Biogeographic regions of Europe 2016, version 1 (EEA 2016). The following bio-geographic regions had no representatives in NDS: Anatolian, Arctic, outside Europe, Steppic, Black Sea. Therefore, rivers from these regions are excluded from presentation.
Geological.typology	Calcerious, Silicious, Organic type of prevailing geology derived from available geological maps	9	W_GEO_CAL W_GEO_SIL W_GEO_ORG W_GEO_OTH (1-0 values)	Geological classification of soil dominating parent material derived from European Soil Database v2.0 (ESDB v 2.0; Panagos 2006) classified into four geological types: calcareous, siliceous, organic, other as presented in Table 2.
Wetted.width	Average river width derived from several measurements at a given site	9	Attribute not considered	
Strahler.stream.order	Stream order according to Strahler (1957) – small streams has	8	W_STRAHLER	Stream order according to Strahler (1957) – small streams has order “1”, when two of them meet – the river has

IC Variable	Description	Importance	New Variable Code	Description
	order “1”, when two of them meet – the river has order “2”, when two rivers of the order 2 met – the order growth to ”3” etc.			order “2”, when two rivers of the order 2 met – the order growth to ”3” etc., derived from CCM2 database.
Geomorph.river.type	Information in 5 categories to be selected, data based on field observations and maps, for river segment including sampled site: Naturally constraint no mob, Braided, Sinuous, Meander regular, Meander tortous.	7	Attribute not considered	
Natural.sediment	Granulation of naturally predominant sediment in 3 classes: 1- coarse (rocks, boulders, stones, gravel), 2 – medium size (sand), 3 – small (mud, clay, peat).	6	Attribute not considered	
E_water_source_type	Predominant source of water: pluvial, pluvio-nival, glacial, groundwaters	5	Attribute not considered	
E_floodplain	Presence of the floodplain (in natural, non-disturbed conditions)	4	Attribute not considered	

The Catchment Characterisation and Modelling River and Catchment Database, version 2.1 (CCM2, Vogt *et al.* 2007) was used to derive physical river and catchment characteristics, i.e. size of catchment, Strahler stream order, slope of the river segment, and mean altitude of a river segment. The Biogeographical regions of Europe 2016, version 1 (EEA 2016) dataset was used to determine the following biogeographical regions: Anatolian, Arctic, outside Europe, Atlantic, Continental, Pannonian, Steppic, Mediterranean, Alpine, Boreal, Black Sea. Finally, the European Soil Database v2.0 (ESDB v 2.0, Panagos 2006) was used to classify the rivers into four geological types: calcareous, siliceous, organic, other). This classification was based on lithological information derived from the database and the details are presented in **Table 2.**

The Following geomorphic descriptors were assigned to the sampling sites:

- Catchment area (km^2) (W_CATCH)
- Valley Slope (‰) (W_Slope)
- Strahler Stream Order (W_STRAHL)
- Mean altitude of the river segment (m a.s.l.) (W_ALT)

- Bio-geographical regions
- Geological type derived from soil dominating parent material (**Table 2**)

The remaining descriptors, i.e. wetted width, geomorphological river type, natural sediment, source of water and presence of floodplain, were either unavailable with the required accuracy for all rivers or were deemed to be too sensitive to human induced alteration (e.g. Geomorphic river type) and were therefore excluded from further analysis (*see* **Table 1** for description).

Table 2. Geological classification based on the European Soil Database v 2.0.

Attributes of European Soil Database v 2.0			Assigned attributes
[STU_SGDBE. MAT11]	**Parent soil material [STU_SGDBE. MAT1]**	**Secondary information on parent soil material [STU_SGDBE.PARMADO]**	**Geological Type**
1 Undifferentia-ted alluvial deposits (or glacial deposits)	100 Undifferentiated alluvial deposits (or glacial deposits)	5000 unconsolidated deposits (alluvium, weathering residuum and slope deposits)	siliceous
		6000 unconsolidated glacial deposits/glacial drift	siliceous
	110 River alluvium	5300 fluvial sands and gravels 5400 fluvial clays, silts and loams	siliceous
	111 Old fluviatile deposit (Tertiary)	5300 fluvial sands and gravels 5400 fluvial clays, silts and loams 5420 river loam	siliceous
	112 Terraces	5310 river terrace sand or gravel 5311 river terrace sand 5400 fluvial clays, silts and loams 5411 terrace clay and silt	siliceous
	113 Lacustrofluvial alluvium	5500 lake deposits	siliceous
	120 Estuarine/Marine alluvium	5100 marine and estuarine sands 5200 marine and estuarine clays and silts	siliceous
	130 Glaciofluvial deposits	6200 glaciofluvial deposits	siliceous
	131 Till	6111 boulder clay	siliceous
	140 Glaciofluvial drift	6200 glaciofluvial deposits	siliceous
	150 Colluvium	5820 colluvial deposit	siliceous
2 Calcareous rocks	200 Calcareous rocks	2100 calcareous rocks	calcareous
	209 Residuum from calcareous rocks		calcareous

Attributes of European Soil Database v 2.0			Assigned attributes
	210 Limestone	2110 limestone	calcareous
	211 Primary limestone (Carboniferous)	2110 limestone	calcareous
	212 Secondary limestone	2110 limestone	calcareous
	213 Tertiary limestone	2110 limestone	calcareous
	214 Ferrugineous limestone		calcareous
	215 Hard limestone	2111 hard limestone	calcareous
	216 Soft limestone	2112 soft limestone	calcareous
	217 Marly limestone	2113 marly limestone	calcareous
	218 Chalky limestone	2114 chalky limestone	calcareous
	219 Detrital limestone	2115 detrital limestone	calcareous
	220 Secondary chalk	2150 chalk	calcareous
	230 Marl	2140 marl	calcareous
	231 Secondary marl	2140 marl	calcareous
	232 Tertiary marl	2140 marl	calcareous
	233 Gypseous marl	2142 gypsiferous marl	calcareous
	234 Schistose marl	2140 marl	calcareous
	240 Gypsum	2210 gypsum	calcareous
	250 Dolomite	2120 dolomite	calcareous
3 Clayey materials – non calcareous	300 Clayey materials	1400 facies bound rock	siliceous
	310 Old clayey sedimentary deposits	1300 pelite, lutite or argillite 5210 pre-quaternary clay and silt	siliceous
	311 Primary clay and sandstone	1300 pelite, lutite or argilite	siliceous
	312 Secondary clay	1310 claystone/mudstone	siliceous
	313 Tertiary clay	5211 tertiary clay	siliceous

Attributes of European Soil Database v 2.0			Assigned attributes
	314 Pleistocene clay	5220 quaternary clay and silt	siliceous
	319 Residuum from old clayey sedimentary deposits	5700 residual and redeposited clays from calcareous rocks	siliceous
	320 Alluvial or glaciofluvial clay	5431 floodplain clay and silt 6111 boulder clay	siliceous
	321 Tertiary alluvial clay	5211 tertiary clay	siliceous
	322 Glacial clay (Tertiary and Quaternary)	6111 boulder clay	siliceous
	323 Gravelly clay	5721 stony clay	siliceous
	324 Boulder clay	5721 stony clay 6111 boulder clay	siliceous
	340 Claystone, mudstone	1310 claystone/mudstone	siliceous
3 Clayey materials - calcareous	330 Residual clay from calcareous rocks	5710 residual clay	calcareous
	331 Clay-with-flints	5711 clay with flints	calcareous
	332 Siderolith formations	5712 ferruginous residual clay	calcareous
	333 Calcareous decalcification clay	5713 calcareous clay	calcareous
	350 Calcareous clay	5713 calcareous clay	calcareous
4 Sandy materials	400 Sandy materials	5311 river terrace sand	siliceous
	410 Old sandy sedimentary deposits	1200 psammite or arenite	siliceous
	411 Secondary sands	1200 psammite or arenite	siliceous
	412 Tertiary sands	1200 psammite or arenite	siliceous
	413 Flint sands	1400 facies bound rock	siliceous
	414 Pleistocene sands	5311 river terrace sand 6210 outwash sand, glacial sand	siliceous

Attributes of European Soil Database v 2.0			Assigned attributes
	419 Residuum from old sandy sedimentary deposits	5600 residual and redeposited loams from silicate rocks	siliceous
	420 Alluvial or glaciofluvial sands	6210 outwash sand, glacial sand	siliceous
	421 Glacial sands	6210 outwash sand, glacial sand	siliceous
	422 Sandy gravelly materials	5300 fluvial sands and gravels	siliceous
	429 Residuum from alluvial or glaciofluvial sands		siliceous
	430 Eolian sands	7200 eolian sands	siliceous
	431 Locally sandcover	7220 cover sand	siliceous
	440 Coastal sands (Dune sands)	7210 dune sand	siliceous
	441 Shelly coastal sands	5121 Holocene coastal sand with shells	siliceous
	442 Non calcareous coastal sands	5100 marine and estuarine sands	siliceous
	450 Sandstone	1210 sandstone	siliceous
	451 Calcareous sandstone (Macigno)	1211 calcareous sandstone	calcareous
	452 Ferrugineous sandstone (Old Red sandstone)	1212 ferruginous sandstone	siliceous
	453 Clayey sandstone	1213 clayey sandstone	siliceous
	454 Soft quartzy sandstone	1210 sandstone	siliceous
	455 Hard quartzy sandstone	1214 quarzitic sandstone/orthoquartzite	siliceous
	456 Quartzite	1214 quarzitic sandstone/orthoquartzite	siliceous
	457 Schistose sandstone	1210 sandstone	siliceous
	459 Residuum	5600 residual and redeposited loams from silicate rocks	siliceous

Attributes of European Soil Database v 2.0			Assigned attributes
	from sandstone		
5 Loamy materials	500 Loamy materials	5000 unconsolidated deposits (alluvium, weathering residuum and slope deposits)	siliceous
	510 Residual loam	5610 residual loam	siliceous
	511 Old loam (Touyas)	5610 residual loam	siliceous
	512 Stony loam	5611 stony loam	siliceous
	513 Clay loam	5612 clayey loam	siliceous
	514 Sandy loam	5610 residual loam	siliceous
	520 Eolian loam	7110 loamy loess	siliceous
	521 Loess	7100 loess	siliceous
	522 Thin loess cover	7100 loess	siliceous
	523 Sandy loess	7120 sandy loess	siliceous
	530 Siltstone	1320 siltstone	siliceous
	539 Residuum from siltstone	5600 residual and redeposited loams from silicate rocks	siliceous
6 Detrital formations	600 Detrital formations	1000 consolidated-clastic-sedimentary rocks	siliceous
	610 Arkose	1220 arkose	siliceous
	620 Breccia and Puddingstone	1110 conglomerate 1120 breccia	siliceous
	630 Flysch and Molasse	1410 flysch 1420 molasse	siliceous
	640 Ranas	1000 consolidated-clastic-sedimentary rocks	siliceous
7 Crystalline rocks and migmatites	700 Crystalline rocks and migmatites	1300 pelite, lutite or argillite	other
	709 Residuum from crystalline rocks and migmatites	5600 residual and redeposited loams from silicate rocks	other
	710 Acid crystalline rocks (and migmatites)	3100 acid to intermediate plutonic rocks	other
	711 Granite	3110 granite	other
	712 Diorite, Quartzodiorite	3130 diorite	other
	719 Residuum from acid	5600 residual and redeposited loams from silicate rocks	other

Attributes of European Soil Database v 2.0			Assigned attributes
	crystalline rocks		
	720 Non acid crystalline rocks (and migmatites)	3200 basic plutonic rocks	other
	721 Syenite		other
	722 Gabbro	3210 gabbro	other
	723 Serpentine	4410 serpentinite	other
	730 Crystalline metamorphic rocks	4200 acid regional metamorphic rocks	other
	731 Gneiss	4240 gneiss	other
	732 Embrechites	4700 tectogenetic metamorphism rocks or cataclysmic metamorphism	other
	739 Residuum from crystalline metamorphic rocks	5600 residual and redeposited loams from silicate rocks	other
	740 Schists	4200 acid regional metamorphic rocks	other
	741 Micaschists	4230 micaschist	other
	742 Slates	4120 slate	other
	743 Shales	4110 (meta-)shale/argillite	other
	744 Calcschists	4520 calcschist, skam	other
	745 Green schists	4310 greenschist	other
	749 Residuum from schists	5600 residual and redeposited loams from silicate rocks	other
	750 Other metamorphic rocks		other
8 Volcanic rocks	800 Volcanic rocks	3400 acid to intermediate volcanic rocks 3700 pyroclastic rocks (tephra)	other
	809 Residuum from volcanic rocks	5600 residual and redeposited loams from silicate rocks	other
	810 Acid volcanic rocks	3400 acid to intermediate volcanic rocks	other
	819 Residuum from acid volcanic rocks		other
	820 Basic volcanic rocks	3500 basic to ultrabasic volcanic rocks	other

Attributes of European Soil Database v 2.0			Assigned attributes
	821 Phonolites		other
	822 Basalt	3510 basalt	other
	823 Andesite	3430 andesite	other
	824 Rhyolite	3410 rhyolite	other
	825 Volcanic tuff	3710 tuff/tuffstone	other
	830 Volcanic slag		other
9 Other rocks	900 Other rocks	3100 acid to intermediate plutonic rocks 3400 acid to intermediate volcanic rocks 3700 pyroclastic rocks (tephra)	other
9 Other rocks	901 Sedimentary rocks	1000 consolidated-clastic-sedimentary rocks	siliceous
	902 Sedimentary, metamorphic and eruptive rocks	3000 igneous rocks	siliceous
9 Other rocks - organic	910 Organic materials	8000 organic materials	organic
No information	No information		other or no information
		3400 acid to intermediate volcanic rocks	other
		4200 acid regional metamorphic rocks	other
		4500 calcareous regional metamorphic rocks	calcareous
		5000 unconsolidated deposits (alluvium, weathering residuum and slope deposits)	
		5100 marine and estuarine sands	siliceous
		5200 marine and estuarine clays and silts	siliceous
		5400 fluvial clays, silts and loams	siliceous
		5600 residual and redeposited loams from silicate rocks	siliceous
		5700 residual and redeposited clays from calcareous rocks	calcareous
		5800 slope deposits	
		6100 morainic deposits	siliceous
		6200 glaciofluvial deposits	siliceous
		7000 eolian deposits	siliceous
		7100 loess	siliceous
		8100 peat (mires)	organic
		8200 slime and ooze deposits	other

2.2 Habitat Use and Tolerance (HUT) guilds

Habitat use guilds were determined by modifying the fish guild classification created for the EFI+ Project (Melcher *et al.* 2007, Holzer 2008, EFI+ Manual 2009, Logez *et al*. 2012), according to which each of the 309 species occurring in European rivers (EFI+ Manual 2009) is ascribed to the following guilds:

1. Intolerant species
2. Tolerant species
3. Benthic species
4. Rheophilic species
5. Lithophilic species
6. Phytophilic species
7. Insectivorous species
8. Omnivorous species

For the purpose of this project eleven HUT guilds were established by combining particular guild characteristics (**Table 3**):

1) Highly rheophilic, intolerant species,

Includes intolerant, rheophilic and lithophilic species. It includes mostly salmonids, some lampreys, bullheads and strongly rheophilic cyprinids.

2) *Rheophilic benthic species, preferring sandy-gravel bottom substrate*

Only benthic and rheophilic species. Neither lithophilic nor omnivorous species are included, regardless of other assignments. These are mainly sturgeons, barbels, gudgeons and some benthic cyprinids.

3) *Rheophilic water column species, preferring sandy-gravel bottom substrate*

Rheophilic and lithophilic species inhabiting the open water column. These are mainly species such as chub, rainbow trout and common minnow.

4) *Limnophilic benthic species of moderate tolerance*

Rheophilic and lithophilic species, excluding those which are benthic or phytophilic, regardless of tolerance and feeding habits. Mainly includes some barbels, loaches and lampreys, as well as flatfish occurring in estuaries.

5) *Limnophilic water column species of moderate tolerance*

Species that are not tolerant, not benthic, not rheophilic and not phytophilic. This is a large guild consisting mainly of various Cyprinids, including some barbels, nases, daces, minnows and roaches. It also includes loaches, Coregonids and some Salmonids of the genus *Oncorhynchus*.

6) *Rheophilic water column species of moderate tolerance*

These are non-benthic species and neither highly tolerant nor intolerant. These are certain daces, red roach, twaite shad and smelt.

7) *Intolerant, water column species*

Species that are intolerant and not benthic such as: Allis shad, bitterling, some Coregonids and Salmonids.

8) *Limnophilic lithophilic species of moderate tolerance*

Lithophilic species that are neither highly tolerant nor intolerant. Also, they are not benthic nor rheophilic nor phytophilic. Mainly species such as asp, Macedonic shad and some daces.

9) Limnophilic phytophilic species of moderate tolerance

Phytophilic species which are neither tolerant nor intolerant, nor rheophilic, nor lithophilic. These are such fish as Siberian sturgeon, some loaches, weatherfish, pike, catfish and rudd.

10) Benthic species of moderate tolerance

Benthic species that are not tolerant. These guild consist of European sturgeon, burbot, some barbels and Neogobiids.

11) Generalists - tolerant species

Only species that are tolerant are chosen regardless of if they belonged to any other guild too. These are mainly cyprinids, such as roach, bleak, common carp, or bream, but also perch and most alien invasive species.

Table 3. Habitat use and tolerance guilds distinguished based on Melcher's guild classification (columns) used for Target Fish Communities development. Key: 1 – species belonging to a guild, 0 – species not belonging to a guild, x – guild not taken into account.

Guild No	Guild	Intolerant	Tolerant	Benthic	Rheophilic	Lithophilic	Phytophilic	Omnivorousus	N species
1	Highly rheophilic, intolerant species	1	0	x	1	1	0	0	20
2	Rheophilic benthic species, preferring sandy-gravel bottom substrate	x	x	1	1	0	x	0	33
3	Rheophilic water column species, preferring sandy-gravel bottom substrate	x	x	0	1	1	0	x	7
4	Limnophilic benthic species of moderate tolerance	x	0	1	0	0	0	0	14
5	Limnophilic water column species of moderate tolerance	x	0	0	0	x	0	x	139
6	Rheophilic water column species of moderate tolerance	0	0	0	x	x	x	x	6
7	Intolerant, water column species	1	0	0	x	x	x	x	12
8	Limnophilic lithophilic species of moderate tolerance	0	0	0	0	1	0	x	9
9	Limnophilic phytophilic species of moderate tolerance	0	0	x	0	0	1	x	23
10	Benthic species of moderate tolerance	x	0	1	x	x	x	x	11
11	Generalists - tolerant species	0	1	x	x	x	x	x	35

The list of species assigned to different guilds is provided in **Table 4.**

Table 4. Guild assignment of European fish species according to the EFI+ Manual and based on Melcher's guild classification.

Species	Guild No	Species	Guild No	Species	Guild No
Alburnoides bipunctatus	1	*Clarius batrachus*	5	*Scardinius racovitzai*	5
Chondrostoma miegii	1	*Cobitis bilineata*	5	*Scardinius scardafa*	5
Cobitis calderoni	1	*Cobitis dalmatina*	5	*Sygnathus abaster*	5

Species	Guild No	Species	Guild No	Species	Guild No
Cottus gobio	1	*Cobitis elongata*	5	*Tilapia zillii*	5
Cottus koshewnikovi	1	*Cobitis elongatoides*	5	*Vimba elongata*	5
Cottus petiti	1	*Cobitis megaspila*	5	*Vimba melanops*	5
Cottus poecilopus	1	*Cobitis narentana*	5	*Zigel balcanicus*	5
Eudontomyzon mariae	1	*Cobitis ohridana*	5	*Zosterisessor ophiocephalus*	5
Hucho	1	*Cobitis rhodopensis*	5	*Alosa fallax*	6
Lampetra fluviatilis	1	*Cobitis tanaitica*	5	*Leuciscus idus*	6
Lampetra planeri	1	*Cobitis vettonica*	5	*Leuciscus pleurobipunctatus*	6
Leuciscus souffia	1	*Cobitis zanandreai*	5	*Osmerus e. eperlanus*	6
Petromyzon marinus	1	*Coregonus autumnalis*	5	*Rutilus arcasii*	6
Salmo salar	1	*Coregonus muscun*	5	*Tropidophoxinellus spartiaticus*	6
Salmo trutta fario	1	*Coregonus oxyrinchus*	5	*Alosa*	7
Salmo trutta	1	*Coregonus pidschian*	5	*Anaecypris hispanica*	7
Salvelinus fontinalis	1	*Coregonus sp*	5	*Aphanius iberus*	7
Thymallus thymallus	1	*Coregonus spp*	5	*Coregonus albula*	7
Zingel streber	1	*Coregonus trybomi*	5	*Coregonus lavaretus*	7
Zingel zingel	1	*Dicentrarchus labrax*	5	*Leuciscus keadicus*	7
Abramis sapa	2	*Dicentrarchus punctatus*	5	*Pungitius hellenicus*	7
Acipenser gueldenstaedtii	2	*Eriocheir sinensis*	5	*Rhodeus sericeus*	7
Acipenser naccarii	2	*Eudontomyzon danfordi*	5	*Salmo trutta lacustris*	7
Acipenser nudiventris	2	*Eudontomyzon stankokaramani*	5	*Salvelinus alpinus*	7
Acipenser ruthenus	2	*Eudontomyzon vladykovi*	5	*Valencia hispanica*	7
Acipenser stellatus	2	*Gobio banarescui*	5	*Valencia letourneuxi*	7
Ambloplites rupestris	2	*Gobio benacensis*	5	*Alosa macedonica*	8
Barbatula barbatula	2	*Hypophthalmichthys nobilis*	5	*Aspius aspius*	8
Barbus barbus	2	*Iberocypris palaciosi*	5	*Chalcalburnus chalcoides*	8
Barbus cyclolepis	2	*Ictalurus nebulosus*	5	*Coregonus peled*	8
Barbus euboicus	2	*Ictalurus punctatus*	5	*Leuciscus borysthenicus*	8
Barbus haasi	2	*Ictiobus niger*	5	*Leuciscus carolitertii*	8
Barbus meridionalis	2	*Knipowitschia goerneri*	5	*Leuciscus pyrenaicus*	8
Barbus peloponnesius	2	*Knipowitschia thessala*	5	*Polyodon spathula*	8
Chondrostoma nasus	2	*Ladigesocypris ghigii*	5	*Thymallus baicalensis*	8
Chondrostoma polylepis	2	*Lepomis auritus*	5	*Acipenser baeri*	9
Chondrostoma toxostoma	2	*Lepomis cyanellus*	5	*Aphanius fasciatus*	9
Chondrostoma vardarense	2	*Lethenteron camtschaticum*	5	*Atherina boyeri*	9
Chondrostoma willkommii	2	*Lethenteron zanandreai*	5	*Clarius gariepinus*	9
Cobitis vardarensis	2	*Leuciscus aradensis*	5	*Cobitis hellenica*	9
Gobio albipinnatus	2	*Leuciscus burdigalensis*	5	*Cobitis taenia*	9
Gobio elimeius	2	*Leuciscus illyricus*	5	*Cobitis trichonica*	9
Gobio gobio	2	*Leuciscus lucumonis*	5	*Economidichthys pygmaeus*	9
Gobio kesslerii	2	*Leuciscus microlepis*	5	*Economidichthys trichonis*	9
Gobio uranoscopus	2	*Leuciscus montenigrinus*	5	*Esox lucius*	9
Gymnocephalus baloni	2	*Leuciscus muticellus*	5	*Eupallasella perenurus*	9
Gymnocephalus schraetser	2	*Leuciscus polylepis*	5	*Knipowitschia caucasica*	9
Huso huso	2	*Leuciscus svallize*	5	*Leucaspius delineatus*	9
Pachychilon pictum	2	*Leuciscus torgalensis*	5	*Misgurnus fossilis*	9
Rutilus pigus	2	*Leuciscus turskyi*	5	*Pseudophoxinus stymphalicus*	9

Species	Guild No	Species	Guild No	Species	Guild No
Sabanejewia balcanica	2	*Leuciscus ukliva*	5	*Sabanejewia aurata*	9
Vimba vimba	2	*Leuciscus zrmanjae*	5	*Scardinius acarnanicus*	9
Zingel asper	2	*Liza aurata*	5	*Scardinius erythrophthalmus*	9
Abramis ballerus	3	*Liza saliens*	5	*Silurus aristotelis*	9
Leuciscus cephalus	3	*Micropterus dolomieui*	5	*Silurus glanis*	9
Leuciscus leuciscus	3	*Misgurnus anguillicaudatus*	5	*Squalius alburnoides*	9
Oncorhynchus kisutch	3	*Morone saxatilis*	5	*Tropidophoxinellus hellenicus*	9
Oncorhynchus mykiss	3	*Mugil cephalus*	5	*Umbra krameri*	9
Phoxinus phoxinus	3	*Mylopharyngodon piceus*	5	*Acipenser sturio*	10
Rutilus frisii	3	*Neogobius melanostomus*	5	*Barbus graecus*	10
Barbus albanicus	4	*Odonthestes bonariensis*	5	*Barbus microcephalus*	10
Cobitis arachthosensis	4	*Oncorhynchus gorbuscha*	5	*Barbus prespensis*	10
Cobitis punctilineata	4	*Oncorhynchus tschawytscha*	5	*Cobitis meridionalis*	10
Cobitis stephanidisi	4	*Orconectes limosus*	5	*Lota lota*	10
Cobitis strumicae	4	*Oreochromis mossambicus*	5	*Neogobius kessleri*	10
Eudontomyzon hellenicus	4	*Oreochromis niloticus*	5	*Proterorhinus marmoratus*	10
Gymnocephalus cernuus	4	*Osmerus eperlanus*	5	*Salaria fluviatilis*	10
Hemichromis bimaculatus	4	*Pachychilon macedonicum*	5	*Salvelinus namaycush*	10
Knipowitschia milleri	4	*Pacifastacus leniusculus*	5	*Triglopsis quadricornis*	10
Knipowitschia panizzae	4	*Parabramis pekinensis*	5	*Abramis brama*	11
Neogobius fluviatilis	4	*Pelecus cultratus*	5	*Alburnus alburnus*	11
Neogobius gymnotrachelus	4	*Perccottus glenii*	5	*Ameiurus melas*	11
Platichthys flesus	4	*Phoxinellus adspersus*	5	*Ameiurus nebulosus*	11
Pleuronectes platessa	4	*Phoxinellus alepidotus*	5	*Anguilla anguilla*	11
Sander volgensis	4	*Phoxinellus croaticus*	5	*Barbus bocagei*	11
Alburnus albidus	5	*Phoxinellus epiroticus*	5	*Barbus comizo*	11
Alosa immaculata	5	*Phoxinellus fontinalis*	5	*Barbus graellsii*	11
Alosa killarnensis	5	*Phoxinellus ghetaldii*	5	*Barbus sclateri*	11
Alosa maeotica	5	*Phoxinellus metohiensis*	5	*Blicca bjoerkna*	11
Alosa tanaica	5	*Phoxinellus prespensis*	5	*Carassius auratus*	11
Alosa vistonica	5	*Phoxinellus pstrossii*	5	*Carassius carassius*	11
Astacus astacus	5	*Poecilia reticulata*	5	*Carassius gibelio*	11
Atherina hepsetus	5	*Pseudophoxinus beoticus*	5	*Chondrostoma lusitanicum*	11
Atherina presbyter	5	*Pseudophoxinus minutus*	5	*Cichlasoma facetum*	11
Aulopyge huegelii	5	*Pungitius platygaster*	5	*Cobitis paludica*	11
Austropotamobius torrentium	5	*Romanichthys valsanicola*	5	*Ctenopharyngodon idella*	11
Barbatula bureschi	5	*Rutilus aula*	5	*Cyprinus carpio*	11
Barbatula pindus	5	*Rutilus basak*	5	*Fundulus heteroclitus*	11
Barbus caninus	5	*Rutilus heckelii*	5	*Gambusia affinis*	11
Barbus guiraonis	5	*Rutilus karamani*	5	*Gambusia holbrooki*	11
Barbus macedonicus	5	*Rutilus lusitanicus*	5	*Gasterosteus aculeatus*	11
Barbus plebejus	5	*Rutilus meidingeri*	5	*Hemichromis fasciatus*	11
Barbus steindachneri	5	*Rutilus ohridanus*	5	*Hypophthalmichthys molitrix*	11
Barbus tyberinus	5	*Rutilus prespensis*	5	*Lepomis gibbosus*	11
Chalcalburnus belvica	5	*Rutilus rubilio*	5	*Liza ramada*	11
Chelon labrosus	5	*Rutilus ylikiensis*	5	*Micropterus salmoides*	11
Chondrostoma arrigonis	5	*Sabanejewia bulgarica*	5	*Perca fluviatilis*	11
Chondrostoma genei	5	*Sabanejewia larvata*	5	*Pimephales promelas*	11

Species	Guild No	Species	Guild No	Species	Guild No
Chondrostoma knerii	5	*Sabanejewia romanica*	5	*Pseudorasbora parva*	11
Chondrostoma lemmingii	5	*Salmo trutta*	5	*Pungitius pungitius*	11
Chondrostoma phoxinus	5	*Salmothymus obtusirostris*	5	*Rutilus macrolepidotus*	11
Chondrostoma prespense	5	*Sander lucioperca*	5	*Rutilus rutilus*	11
Chondrostoma soetta	5	*Scardinius graecus*	5	*Tinca tinca*	11
				Umbra pygmaea	11

2.3 Statistical analysis

River water bodies were classified according to the fish assemblages at non-disturbed (reference) sites expected in specific geomorphological settings. Following this, physical attributes that distinguish the type of site expanded this classification to all water bodies in the European database. We applied non-hierarchical clustering (NHC) analysis and discrimination with Analysis of Group Similarities (ANOSIM) for step one, and Classification and Regression Trees (CART, Breiman *et al.* 1984, Clarke 1993, De'ath & Fabricius 2000) for step two.

2.3.1 Cluster Analysis and ANOSIM

Cluster analysis was applied to two data sets sequentially. The first data set consisted of geomorphic descriptors of sampling sites, and was clustered into samples with similar habitat characteristics. These cluster groupings were then added as an additional variable to the biological data of HUT fish guild proportions captured at each site in order to produce a mixed data set (guilds/physical clusters) that was itself then clustered. In both cases, the clustering procedure was the same: a distance matrix was created by standardizing the data using the *Gower* and *Manhattan* similarity distances for the physical and the mixed data set, respectively (Gower 1971, Krause 1987). The number of clusters was determined with the help of scree and silhouette plots. A Partitioning Around Medoids (PAM) clustering model was applied (Kaufman & Rousseeuw 1987, Hastie *et al.* 2001, Park & Jun 2009). The cluster plots and silhouette plots were created, as well as box plots for each variable. Subsequently, data discrimination was performed via ANOSIM in order to verify model performance. The calculated FCMacHT classes were then assigned to each site.

The physical data set with the assigned FCMacHT classes were analysed via CART analysis using the "gini" split method and 30 times smoothing of the CP-Plot (Qian 2016). Pruning was performed manually to arrive at the simplest and most parsimonious solution.

2.4 Map Creation

The map containing all rivers classified into FCMacHT was created using ArcGIS 9.2.1. The Classification and Regression Tree (CART) calculated for non-disturbed sites (section 2.3.1) was used to calculate the FCMacHT for both non-disturbed and impacted rivers, based on the selected physical variables (see **Table 2**).The map presenting rivers classified according to FCMacHT was produced using the surface water bodies' map of the CCM2 dataset.

2.5 Expected Fish Communities

The Target Fish Communities (TFC) method (Bain & Meixler 2008) was used to calculate expected HUT guilds proportions for each of FCMacHT class. However, the data from 1099 NDS underwent critical review and selection process. First, sites with extreme values (<5%) for environmental parameters of river slope, altitude or catchment area were excluded. Next, sites that did not have an International Reference Site status were excluded from the low anthropogenic disturbance category.

An additional 182 sites from Poland were selected as undisturbed during the auto-intercalibration process with the same criteria used for IRS in the Intercalibration database (Prus *et al*. 2016). From the IRS set, 10 representative sites were randomly selected for each FCMacHT class. The random selection was constrained by requirements of equal geographic distribution to avoid over-representation of bio-geographical regions. For each group of 10 sites the sum of guild proportions was calculated and ranked. Then, the reciprocal rank was calculated (1 divided by rank score, so proportionally fewer species had lower values). Thus, by dividing by the sum of the reciprocal ranks one can get the expected proportion of the guild. The proportions were represented by pie charts for each FCMacHT.

3 RESULTS

3.1 Data

The European Intercalibration database consists of 19 tables with data and metadata for 4561 sites. Fish samples were taken with electrofishing between 1958 and 2008. The following countries are represented: Austria, Belgium – Flanders and Wallonia, Czech Republic, Germany, Denmark, Spain, Estonia, United Kingdom (England and Wales, Scotland, Northern Ireland), Finland, France, Greece, Ireland, Lithuania, Luxembourg, Latvia, Netherlands, Norway, Portugal, Romania, Sweden, Slovenia and Slovakia. This database was complemented with the data from Poland (938 fished sites sampled between 2011 and 2015) so the total dataset available was 5499 fished sites, including 1099 NDS.

There was a requirement for individual permissions for data use, which caused some difficulties. The majority of the EU countries (19) granted permission. Germany, France, Denmark and Greece did not respond to data use requests, and thus data from those countries were used only to type European rivers, but not for the computation of expected fish communities. Romania did not grant permission to use its data within the AMBER project and thus was excluded from analyses.

3.2 Statistical analysis

3.2.1 Cluster Analysis

Figure 1 presents scree and silhouette plots for the physical site descriptors, from which 12 river type clusters were selected.

Figure 2 shows the cluster plot for two principal components and **Figure 3** a silhouette plot. The first two components explain 94% of the variability; average silhouette width was 0.48. The ANOSIM discrimination carried out on the resulting clusters is highly significant (R=0.907, P<0.001; **Figure 4**).

Figure 5 shows the scree and silhouette plots for the biological descriptors associated with the physical clusters from **Figure 1**. Based on this figure, we selected eight clusters for the NHC procedure.

Figure 6 shows the cluster plot with the FCMacHT clusters, which appear to be well separated in the 2-D space. However, this separation explains only 13 % of the variability, indicating that other components are important.

Figure 7 shows the silhouette plot with an average silhouette width of 0.53.

Figure 8 shows how individual variables were distributed among classes. It is apparent that the environmental clusters are distributed among FCMacHT clusters almost without overlap. The ANOSIM discrimination was highly significant (R=0.937, P<0.001; **Figure 9)**.
After inspection of the results, we decided to split cluster 8 into two additional clusters, as the results seemed geographically inconsistent and included rivers in Mediterranean mountains as well as rivers in boreal lowlands. Accordingly, Mediterranean mountain rivers were allocated to cluster 8, and Boreal lowland rivers were allocated to cluster 9.

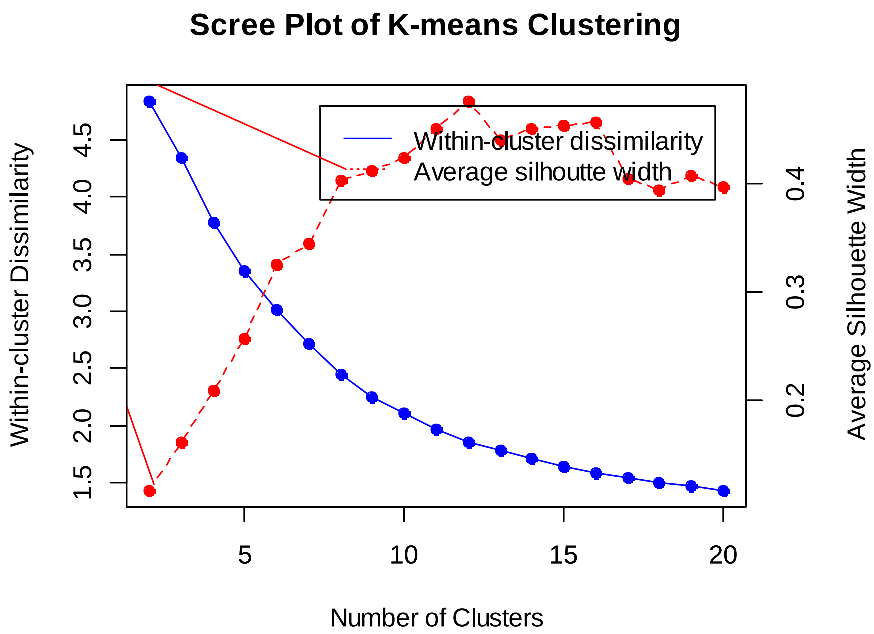

Figure 1. Scree and silhouette plots of environmental attributes of study sites.

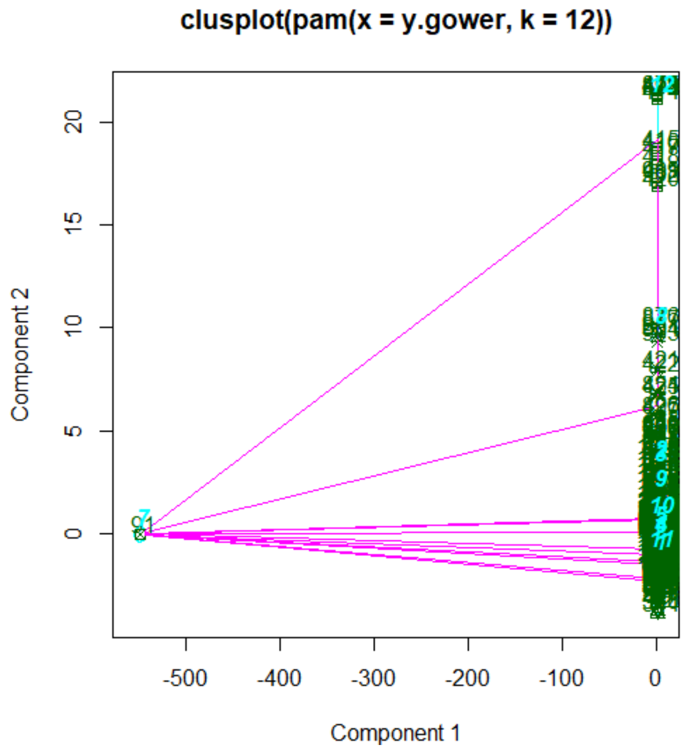

Figure 2. Cluster plot of environmental attributes of study sites for two main components. The numbers represent the site number and the pink lines the distance between centroids.

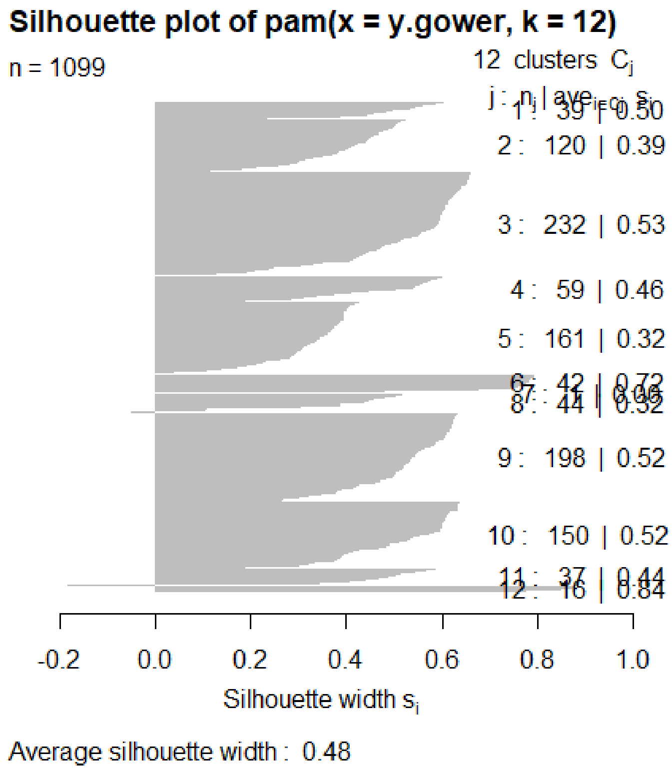

Figure 3. Silhouette plot for clusters of environmental attributes of study sites.

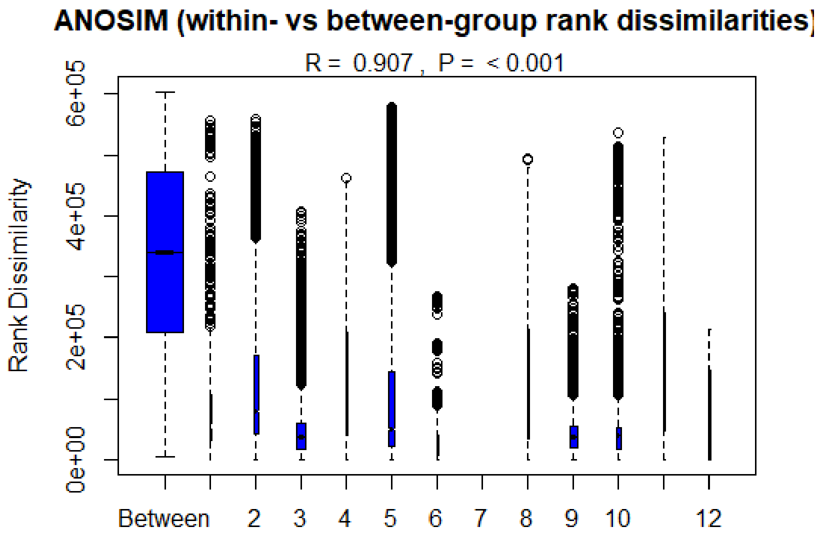

Figure 4. Box plots representing distances between and within clusters of environmental attributes study sites.

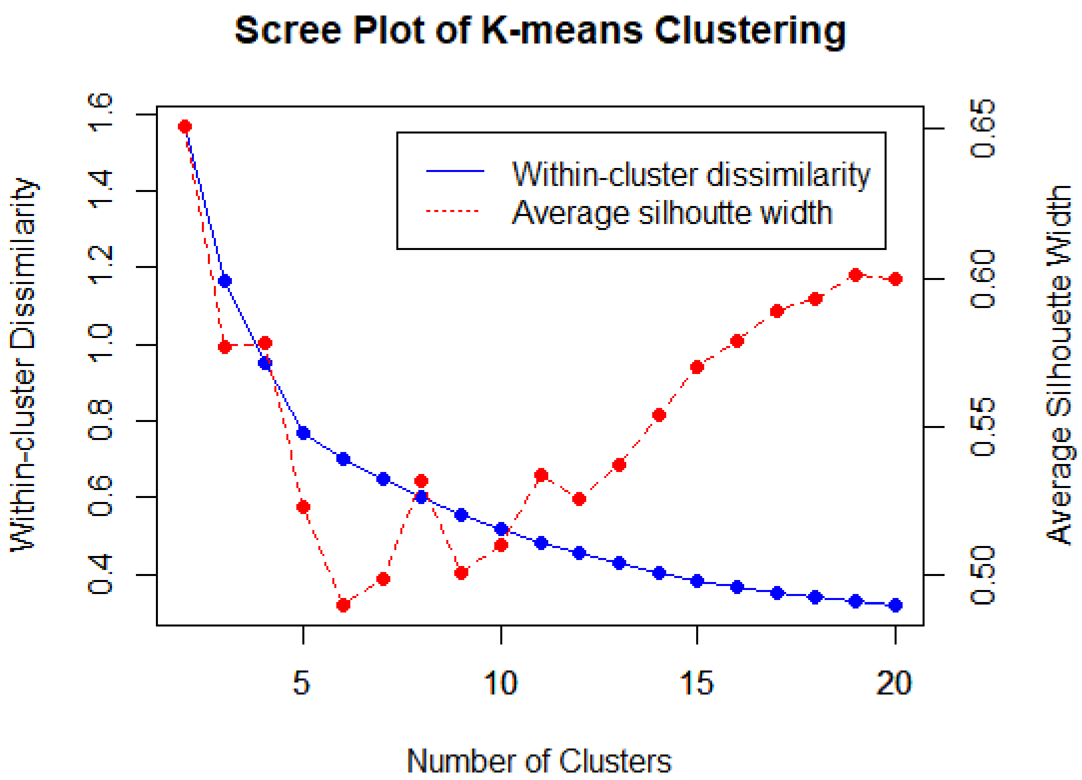

Figure 5. Scree and silhouette plots of habitat attributes (environmental class and guild composition) of study sites.

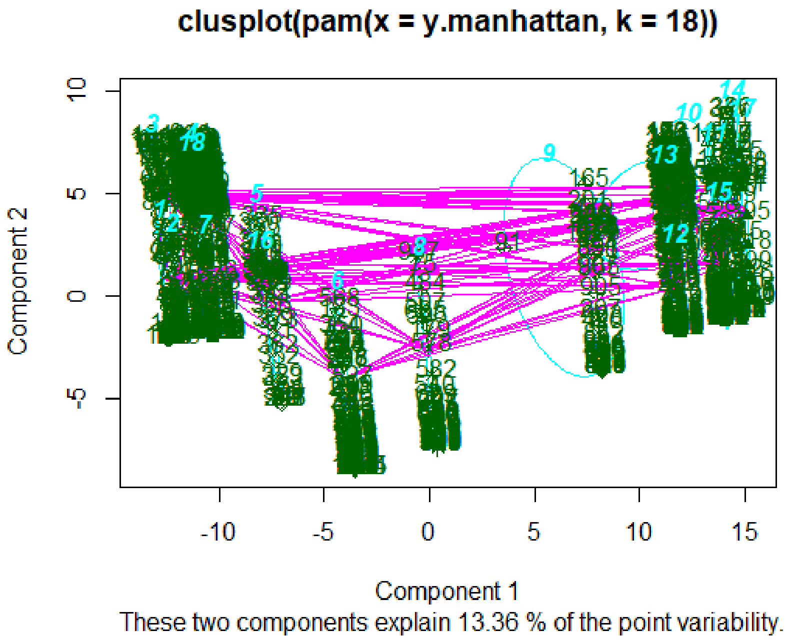

Figure 6. Cluster plot of habitat attributes of study sites for the two main components. The numbers represent the site number and the pink lines the distance between the centroids.

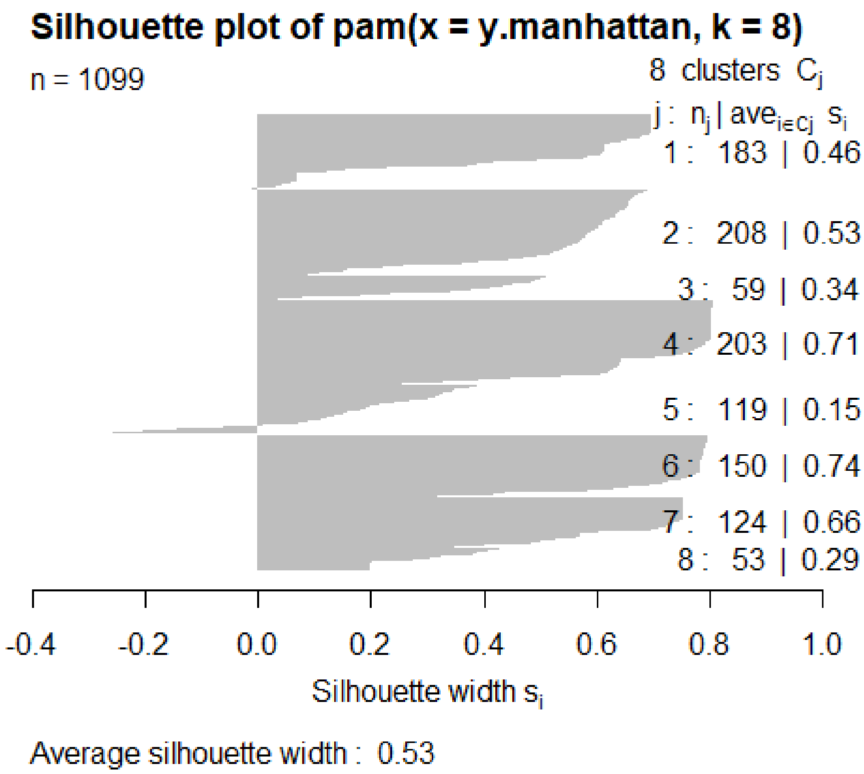

Figure 7. Silhouette plot for clusters of habitat attributes of study sites.

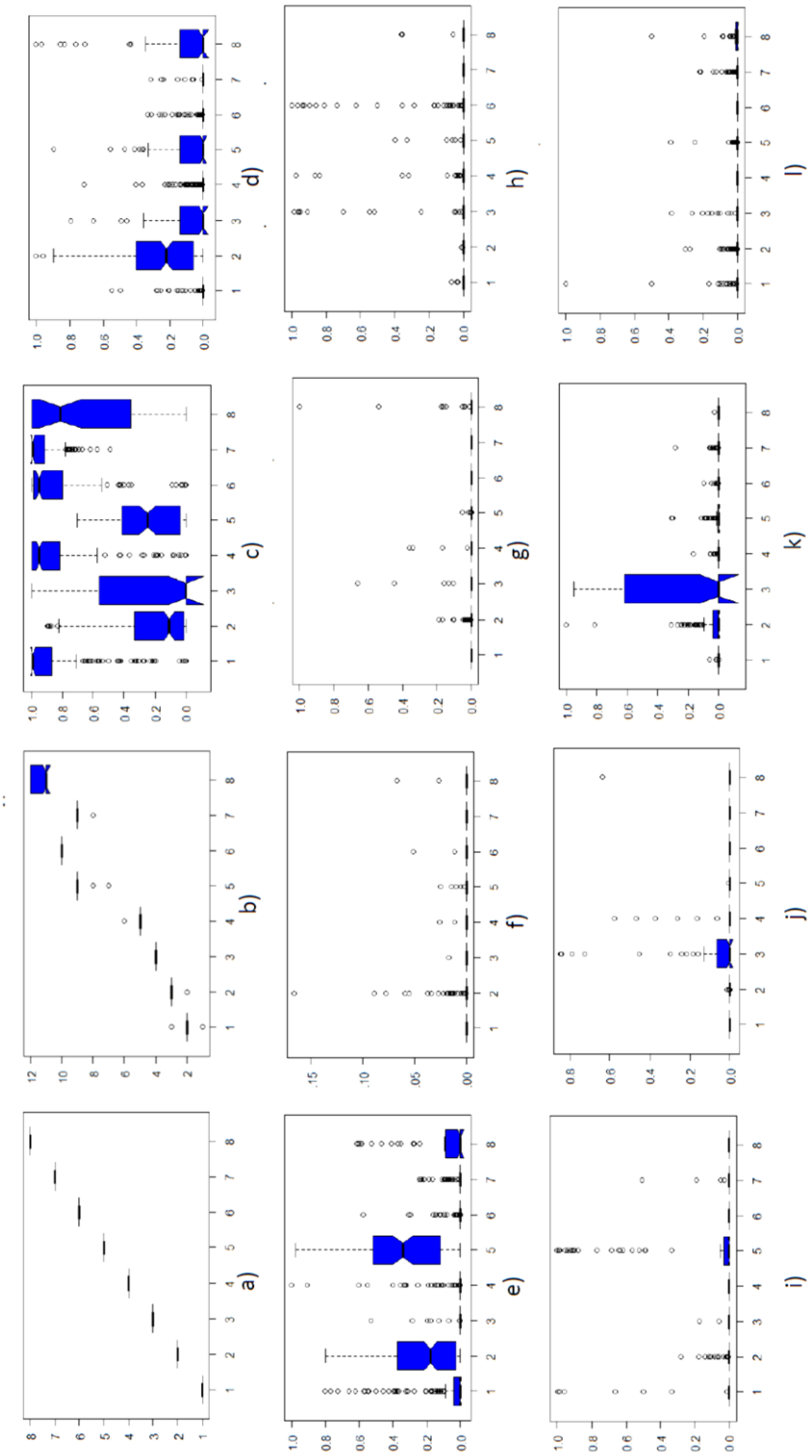

Figure 8. Box plots of distribution of individual parameters among the clusters: a) FCMacHT, b) HYMO class, c) Highly rheophilic, intolerant species, d) Rheophilic benthic species, preferring sandy-gravel bottom substrate, e) Limnophilic benthic species of moderate tolerance, f) Limnophilic water column species of moderate tolerance , g) Rheophilic water column species of moderate tolerance, h) Intolerant, water column species, i) Limnophilic lithophilic species of moderate tolerance, j) Limnophilic phytophilic species of moderate tolerance, k) Benthic species of moderate tolerance.

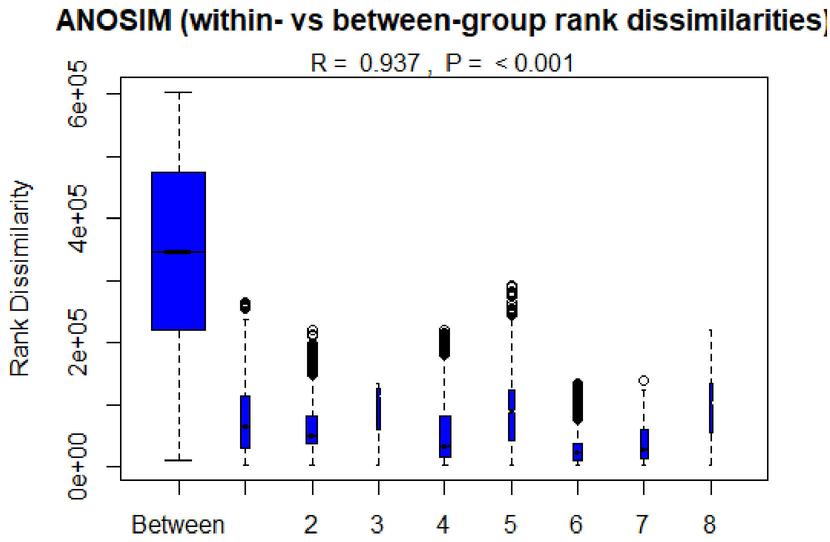

Figure 9. Box plots representing distances between and within the clusters of habitat attributes of study sites.

3.2.2 CART

Figure 10 presents the Complexity Parameter plot and **Figure 11** the Classification tree, which was pruned at 0.003 relative error. The 22 leaves long tree has 94% correct reclassification (Kappa=0.912). The primary splitting variable is Atlantic bio-geographical region, then further Alpine and Continental. Within the Atlantic region there are only two FCMacHT classes 4 and 6 distinguished by their geology. In Boreal region 5 and 7 FCMacHT classes are dominating. Occurrence of entire sample of class 9 in this group confirms earlier decision of separating it from type 8. It is distinguished by low slope and large catchment area. The Types 5 and 7 are separated by various combinations of altitude and catchment size. In Continental region rivers (FCMacHT 1,2,4) are separated by slope and siliceous geology. In other regions rivers are separated by calcareous geology (FCMacHT 3, 8).

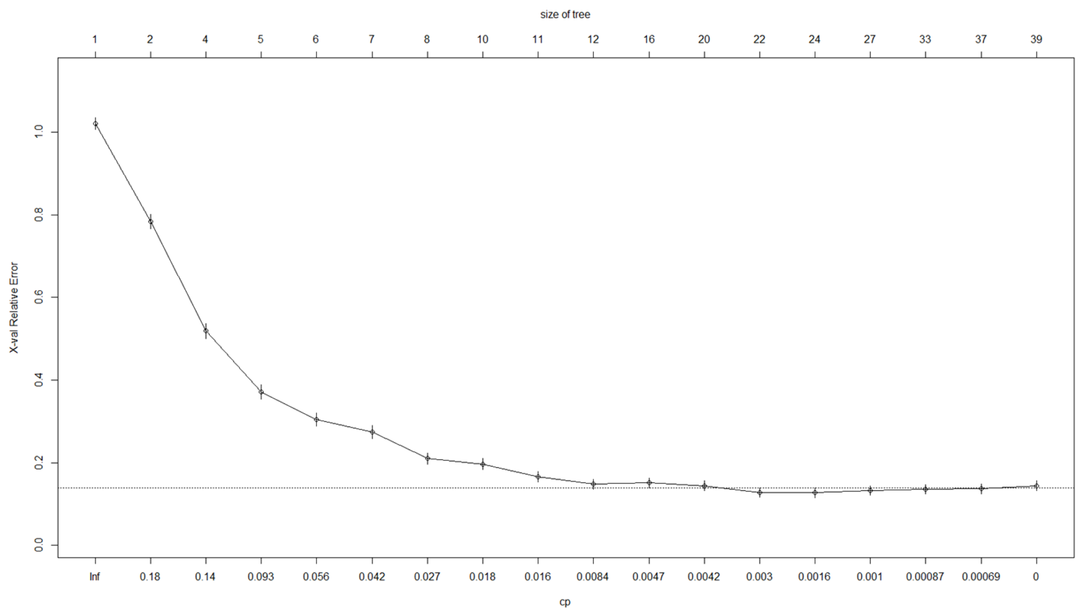

Figure 10. Complexity Parameter plot for environmental attributes of study sites.

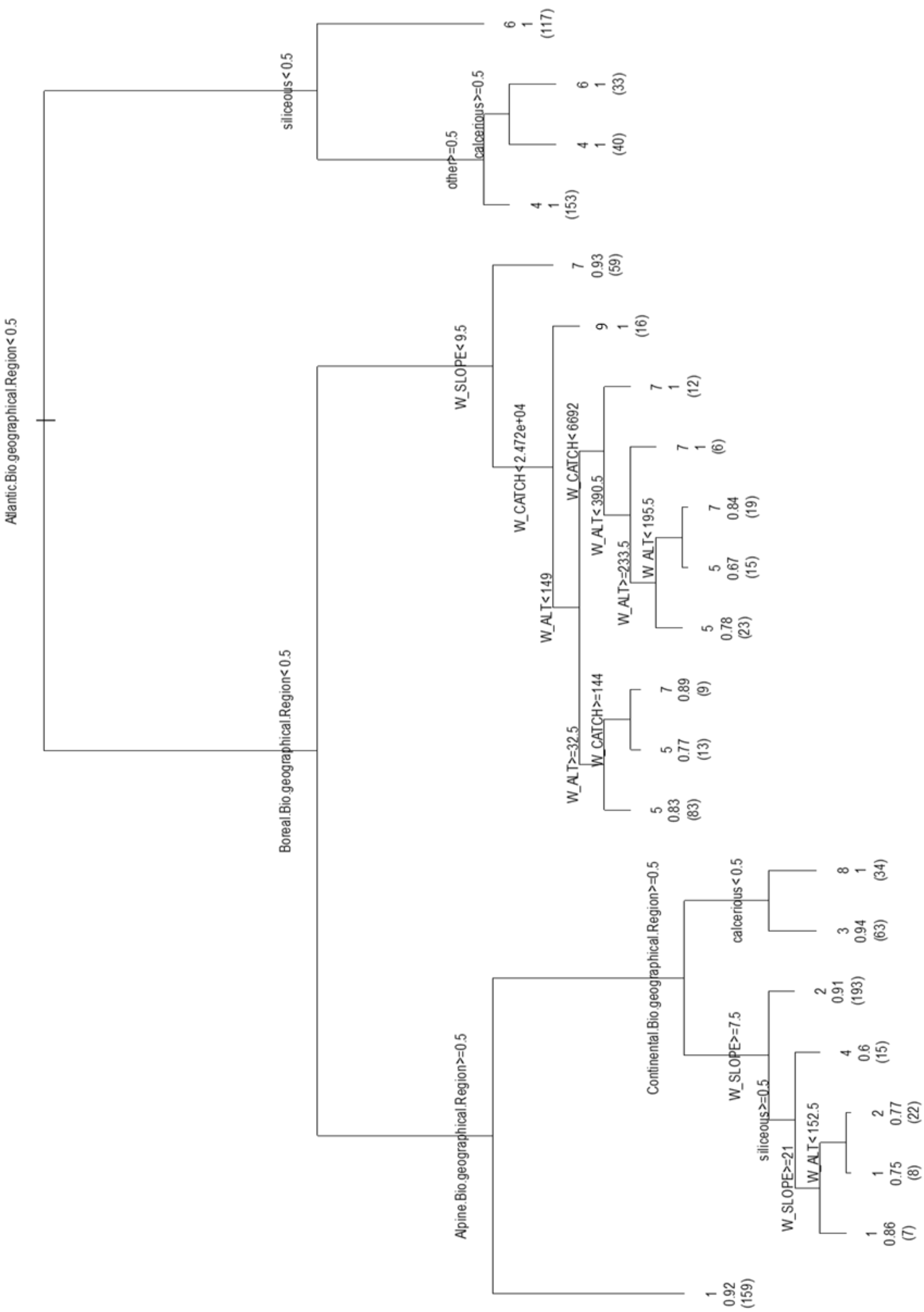

Figure 11. Classification tree of environmental attributes according to the FCMacHT class.

3.3 Map

Figure 12 represents the distribution of rivers classified according to FCMacHT sampled in the Intercalibration exercise.

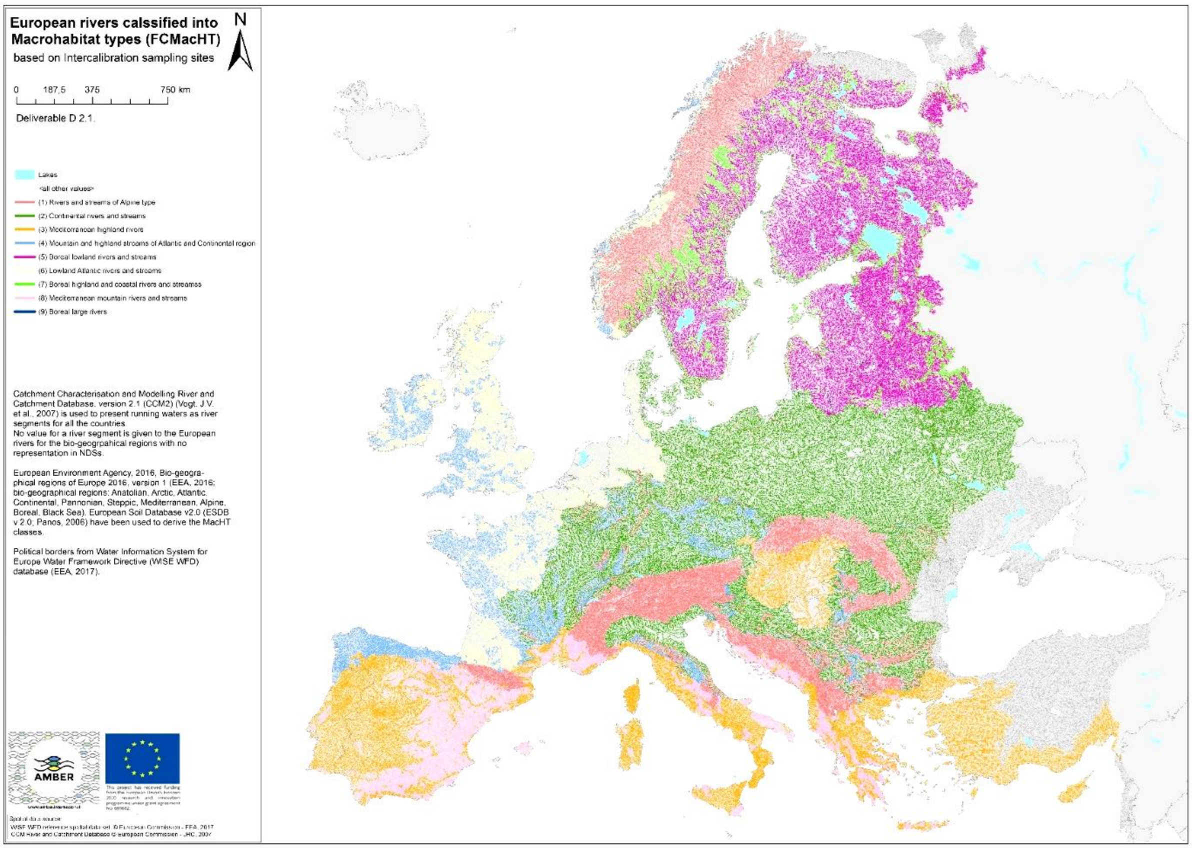

Figure 12. European rivers classified into Macrohabitat types (FCMacHT).

3.4 Expected Fish Communities

An Expected Fish Community (EFC) was developed using the guild proportions in fish assemblages, calculated for a set of 10 IRS sites for each type (**Figure 13**). Basic abiotic and geographic characteristics of those IRS sites chosen for EFC development are described for each river type, together with the expected fish guild composition.

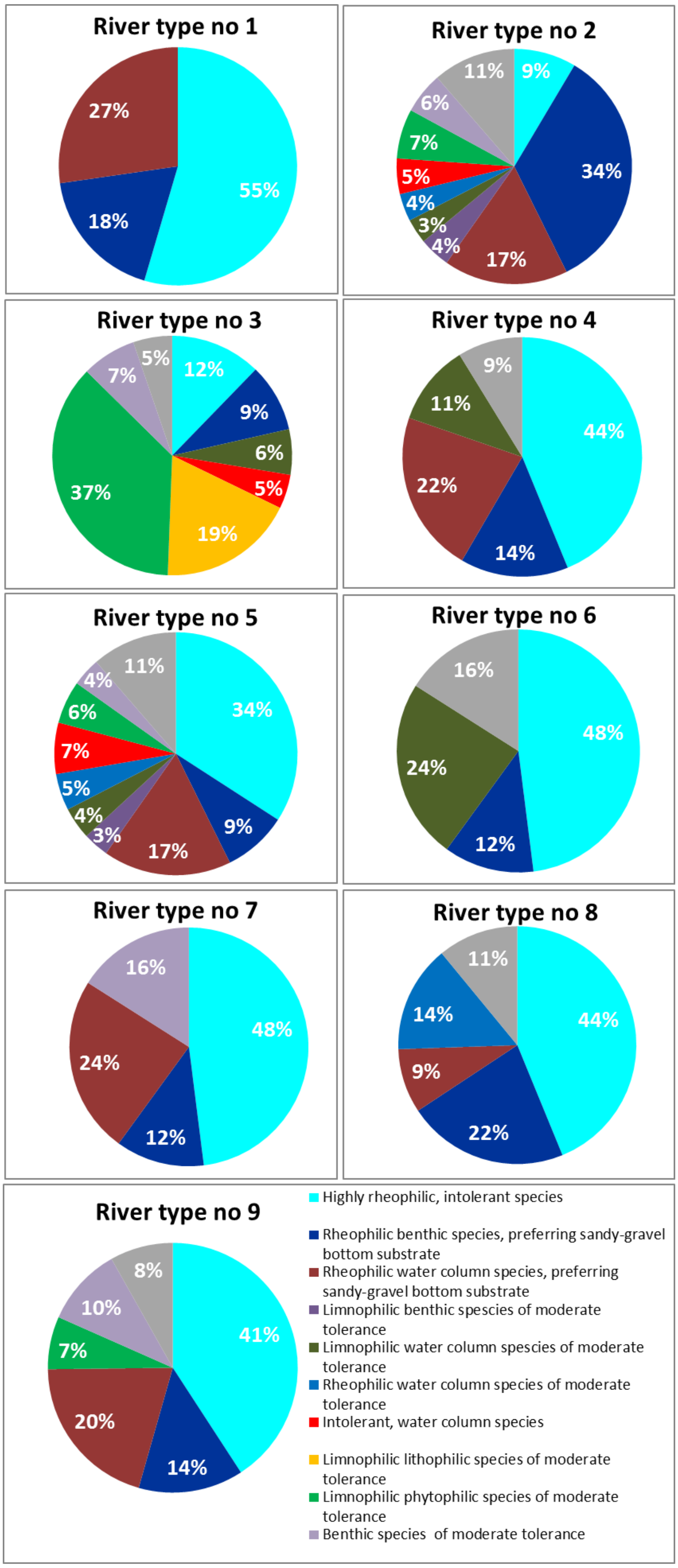

Figure 13. Expected fish guild proportions in 9 European FCMacHT classes. Within each graph, fish guilds are ordered from more rheophilic and intolerant to generalists – tolerant species.

Type 1

The first river type (Type no 1, **Figure 13**) groups rivers of Alpine type located in mountain areas. IRS sites are located up to 934 m a.s.l. (average 626 m a.s.l.) and characterized by high slope values - up to 94‰ (average 29‰). These rivers are distributed across Central and Southern Europe. Fish communities of this type consist mainly of highly rheophilic, intolerant species (55%), with a share of 27% of Rheophilic water column species and 18% of Rheophilic benthic species, preferring sandy-gravelly bottom substrates. Such composition of fish assemblage is associated with rivers that have steep slopes, cold temperatures, well aerated waters, and coarse bottom substrates.

Type 2

River type no 2 can be considered as Continental rivers and streams. IRS sites are located up to 288 m a.s.l. (average 179 m a.s.l.) and characterized by low slope - up to 11‰ (average 3‰). These rivers are distributed across Central Europe – from the Czech Republic, through Poland to Belgium, in the Continental bio-geographical region. The fish community in these rivers (**Figure 13**) is composed of 34% of Rheophilic benthic species, preferring sandy-gravelly substrates. Four rheophilic fish guilds make up 64% of the fish community, while three limnophilic guilds make up 14%. Benthic species of moderate tolerance represent 6%, while Intolerant water column species – 5%. Generalists are moderately abundant with a 11% share of the fish community. The high diversity of guilds reflects the high habitat variability of these rivers which are widely distributed across the Continental bio-geographical region.

Type 3

In the river type no 3 a set of Mediterranean highland rivers is grouped. The maximum height of IRS site is 556 m a.s.l. with an average of 369 m a.s.l. River slopes are quite high – up to 27‰ (average 14‰). These rivers are located mostly in the Iberian Peninsula. Limnophilic phytophilic species of moderate tolerance dominate with a share of 37%. The next group is a guild of Limnophilic-lithophilic species of moderate tolerance (19%), which is observed exclusively in this river type. Two guilds of rheophilic fish represent together 21%. Benthic species of moderate tolerance represent 7% of fish community. Intolerant water column species – 5% share of Generalists in fish community is low – 5% (Figure 13). The high incidence of limnophilic guilds, which are tolerant of high temperatures and low oxygen concentrations, is associated with the Mediterranean climate.

Type 4

In the FCMacHT 4 a set of mountain and highland streams of Atlantic and Continental region is grouped. For IRS sites altitude reaches up to 1050 m a.s.l. (average 497 m a.s.l.). Slope values are high, reaching 72‰ with an average of 29‰. Rivers of this type are located across Europe, from Poland to Spain and from Scotland and Ireland to Slovenia. These streams are characterized with small catchment sizes, reaching up to 122 km2, with an average of 60 km2. The fish community is composed of five guilds, with a strong dominance of rheophilic species (together making 80%), with a dominance of Highly rheophilic, intolerant species (44%) (Figure 13). Limnophilic water column species of moderate tolerance represent 11%, generalists are moderately abundant – 9%. The fish assemblage corresponds to mountainous streams, characterized by high slopes and coarse substrates.

Type 5

Type no 5 groups lowland Boreal rivers and streams. The chosen IRS sites are located on altitudes up to 184 m a.s.l. (average 69 m a.s.l.) with extremely low slope values – maximum of 4‰, and an average of 2‰. Sites used for EFC development are located in Central and Northern Europe, mainly in the Boreal bio-geographical region. This river type has a highly diversified fish assemblage: four guilds of rheophilic species make up 65% of the fish community, with a dominance of highly rheophilic intolerant species (34%), while three guilds of limnophilic species have a 13% share. Intolerant water column species represent 7%, while Benthic species of moderate tolerance – 4%. Generalists are present with a share of 11% (**Figure 13**). The fish community of this river type is complex with almost all guilds present (10 of 11). This reflects the high habitat variability in undisturbed lowland rivers of moderate size. Also a substantial share of rheophilic species confirms the good ecological status of the chosen IRS sites.

Type 6

The No 6 FCMacHT river type groups lowland Atlantic-type rivers and streams. The maximum altitude reaches 180 m a.s.l. (average 66 m a.s.l.) and rivers have moderate slopes, reaching 30‰, with an average of 17‰. Streams used for EFC development are located mainly in Ireland and Scotland, and are characterized with small catchment size – up to 57 km2. The dominant group of fish are highly rheophilic, intolerant species (48%), while Rheophilic benthic species, preferring sandy-gravelly substrates make up 12% of the fish assemblage. The next fish guild was Limnophilic water column species of moderate tolerance – 24%. Generalists were present with a share of 16% (**Figure 13**). The fish community is characteristic of small lowland streams of steep slopes, located in hilly regions near the coast.

Type 7

The No. 7 river type is composed of Boreal highland and coastal rivers and streams. The maximum altitude for IRS sites is 232 m a.s.l. with an average of 132 m a.s.l. IRS sites of this type are distributed in northern Europe: Finland, Sweden and Lithuania. These rivers are characterized by moderate slopes, not exceeding 18‰ (average 5‰). However, rheophilic guilds strongly dominate (three guilds making up to 84% of the fish assemblage). The rest of the fish community is composed of Benthic species of moderate tolerance (16%) (**Figure 13**). Such a strong dominance of rheophilic species is largely determined by the climatic conditions of the Boreal bio-geographical region, characterized by low summer temperatures and high oxygen concentrations.

Type 8

The No. 8 river type is composed of Mediterranean mountain rivers and streams. For IRS sites the altitude reaches up to 1218 m a.s.l. (average 850 m a.s.l.) and slope values are high, reaching 50‰ with an average of 25‰. Rivers of this type are located in Southern Europe. The fish community is composed of four guilds of rheophilic species (together 89%), with a dominance of highly rheophilic, intolerant species (44%). The rest of fish assemblage is formed by generalist – tolerant species (11%) (**Figure 13**). The strong dominance of rheophilic species is associated with the mountainous character of these rivers, while the presence of generalists may be the result of localised, mild Mediterranean climatic conditions.

Type 9

The last river type (no 9) is composed of large Boreal rivers (catchment size up to 40112 km^2) with a maximum altitude of 109 m a.s.l. (average 50 m a.s.l.) and low slopes – up to 3‰ (average 2‰). IRS sites are located in Finland and Sweden. The dominant fish group here are Highly rheophilic, intolerant species (41%) followed by two rheophilic groups that make up 34% of the fish community. Benthic species of moderate tolerance constitute 10% of fish community, a share of Limnophilic phytophilic species is 7%, while that of generalists is 8% (**Figure 13**).

4 DISCUSSION & CONCLUSIONS

4.1 Data used

The obtained intercalibration data represents a valuable cross section of fish habitats occurring in European rivers and illustrates the status of fish communities. We were able to benefit from the wealth of expertise and experience of environmental scientists, who not only provided the data, but also identified habitat attributes expected to correspond with species composition. In the preliminary work for this deliverable we developed predictive models using the variables from Intercalibration data set (not presented here). The performance of these models was similar to that created using the attributes from the Geographical Databases. In subsequent steps presented here we reduced the set of geomorphic variables to those that can be easily found in all the databases and added only Strahler's Stream Order. This metric is frequently used as an important predictor of species distribution in rivers (Kuechne 1962, Matthews & Marsh-Matthews 2017, Strahler 1957). It was also included in the development of the Target Community approach, which is utilized here for determination of Expected Fish Communities associated with FCMacHT classes. In North America the Target Fish Community has been used to predict fish distributions from rivers selected according to Ecoregion, Strahler Order, altitude class and % calcareous geology (Omernik & Griffith 2014). This approach was adopted in this study for Europe, with similar but not identical environmental variables. Our model focuses on fish abundance as an indicator of ecological status, as recommended by the EC-funded MARS project (http://www.mars-project.eu/)

The selected geomorphic attributes can be described as those that are not directly modifiable by human actions and therefore allow us to predict the expected fish community structure as determined by macrohabitat settings. It is the same approach applied by the USGS for the entire Laurentian Great Lakes Basin (McKenna *et al.* 2015), which also used environmental attributes similar to those selected in this project. Most notably though, temperature metrics were not utilized as they were not available. Similarly, Park *et al.* (2005a), in modelling fish community structure with ANN (Artificial Neural Networks) in France identified eight environmental attributes (gradient, elevation, distance from source, surface area, width, depth, July air temperature and January air temperature). In our model distance from source was replaced by Stream Order and average depth is omitted as it as a function of river flow.

The Fish Community Macrohabitat typology presented here is aligned with typologies used in the implementation of WFD as it builds directly upon the classification schemes and data from the Intercalibration exercise. Non-Disturbed Sites (NDS) and International Reference Sites (IRS) are extracted from the Intercalibration procedures and databases. Those reference sites were chosen according to commonly accepted criteria during the Intercalibration process and applied here as a foundation for the model. FCMacHT offers an additional, novel component to the description of reference conditions focusing on macrohabitat distribution as it affects fish community composition.

The key difficulty faced was obtaining permissions from the data owners to use the information in AMBER, despite the fact that the data had been collected as part of EU funded research and should have been by now publicly and easily available. The issue was brought up into the ECOSTAT Commission and triggered appropriate actions towards common consent for use of the data for research purposes, which will help to avoid such delays in the future. Our project highlighted the importance of ensuring that data use agreements are not unnecessarily restrictive, and also the huge importance of data reuse for research and river management purposes. The integrated statistical analysis reported here is a new tool that offers broader insights into river ecology and paves the way for assessing barrier impacts and other human induced alterations that can affect ecological status in the context of the WFD.

4.2 Habitat use and Tolerance guilds

Grouping fish species into HUT guilds is an effective method of generalization of environmental analysis for application at regional and continental scales (Parasiewicz *et al.* 2012, Welcomme *et al.* 2006). It allows the creation of robust models that are not affected by local species variability.

The AMBER project was able to take advantage of earlier accomplishments of European research projects and particularly EFI+ work. The guild classification prepared by the team of scientists under the leadership of Dr. Andreas Melcher served as a solid foundation for establishing of HUT guilds (Melcher *et al.* 2007). Furthermore, defined HUT guilds corroborate those identified independently in Poland during a national research project that developed methodologies for environmental flow regulation (Polish National Water Management Authority 2015).

4.3 Statistical analysis

The use of clustering and multivariate statistics to establish relationships between physical attributes and fish distribution is best described in Park *et al.* (2005b). It presents a number of methods implemented here with varying success. The methodology applied here was chosen for its simplicity and resulted in a high level of explanatory power.

As described by Park *et al.* (2005b), two-step approaches consisting of clustering and discriminant analysis are frequently used for predicting community composition from environmental data. The first clustering step is usually applied to the biological data. However, this approach has the disadvantage of being affected by high level of biological variability and noise as biological assemblages are known to fluctuate rapidly over time. In contrast, environmental attributes such as ecoregions, watershed area or slope are relatively constant. Therefore, we reversed the process and conducted a first clustering step on environmental attributes. In the second step we related these groups to biological data with clustering and discriminant analysis. As a result FCMacHT clusters were distinct even when plotted with only two orthogonal principal components, and discrimination occurred at R=96, which for environmental data represents a strong level of differentiation. Consequently, a follow up CART decision-tree also performed very well, achieving 94% correct classification.

The habitat description of FCMacHT classes appeared logical and intuitive, and could generally be described with simple descriptive names. The community structure and distribution of rivers on the maps appear to make intuitive sense at a large scale, suggesting that our models provide a reasonable and plausible interpretation of the data.

The expected fish communities demonstrate interesting patterns. Among the nine FCMacHT classes **five groups** with similar Expected Fish Community structure are apparent: **1)** Alpine and mountain rivers with communities composed mainly of rheophilic species (types: 1, 4, 8); **2)** Continental rivers – with a dominance of rheophilic benthic species and high diversity of fish guilds - from rheophilic to limnophilic species and generalists (type 2); **3)** Mediterranean rivers – with prevalence of phytophilic and limnophilic species and some share of rheophilic species, benthic species and generalists (type 3); **4)** Atlantic lowland rivers – with dominance of rheophilic species and substantial share of limnophilic species and generalists (type 6); **5)** Boreal rivers – with a high share of rheophilic species and presence of benthic species (types: 5, 7, 9). Hence, in most cases the rheophilic fish guilds prevailed, with the exception of the Mediterranean rivers.

4.4 Uncertainty of spatial data

The base map dataset's accuracy (ECRINS, CCM2 and WISE WFD) is limited with regards to the sampling sites. None of base river maps available include all of the smallest rivers used for sampling in the Intercalibration exercise. The CCM2 dataset seems to be more detailed than the WISE WFD, however, it is less precise with regard to location and spatial extent. On the other hand, the WISE WFD network seems to be more precise for location, but the smallest rivers are not included (**Figure 14**). Additionally, this dataset is not topological.

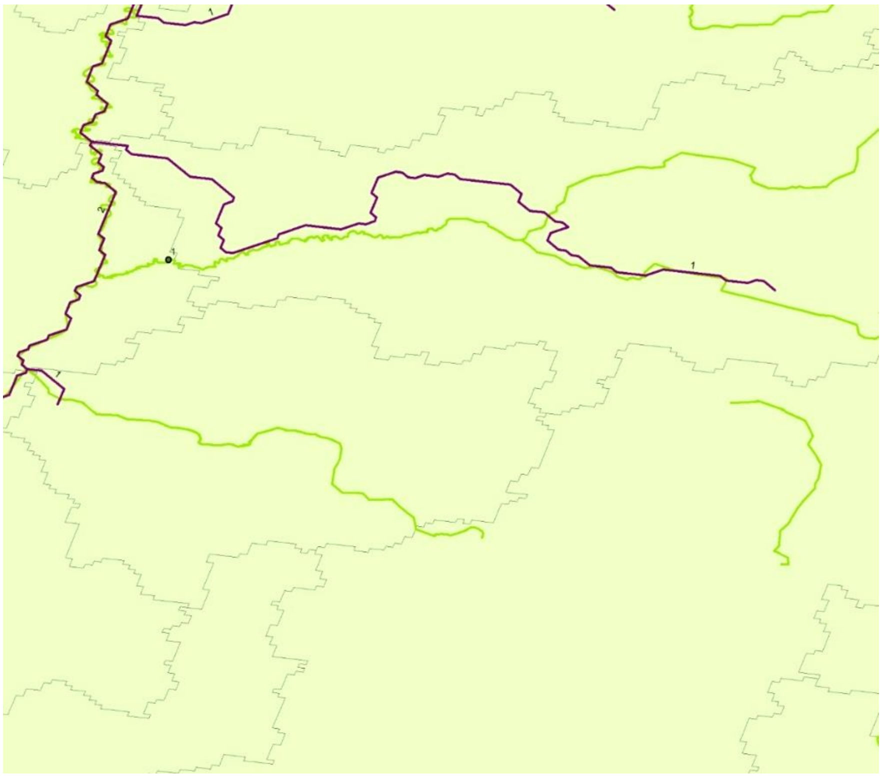

Figure 14. Examples of WISE WFD and CCM2 discrepancies. Green rivers: WISE WFD dataset; purple rivers: CCM2 dataset; grey rivers: catchment boundaries of the CCM2 dataset. Points represent sampling sites from the Intercalibration exercise.

4.4.1 Map presentation and limitations

The WISE WFD spatial reference dataset, which is recommended by the EU for implementation of the WFD, does not include all the European countries; also data quality is not homogenous across countries. Therefore, the CCM2 dataset was used to present all the river classes.

Out of the 1315 NDSs, the CART classification used 1099 non-disturbed sites to calculate macrohabitats groups. Some of these sites (about 20 points) could have been incorrectly classified into siliceous instead of calcareous geology types (and vice versa) due to problems with the interpretation of the geological data. This may have introduced some errors into the CART decision trees, but it is thought that this problem is a minor one.

The CCM2 database structure (see Vogt *et al.* 2007) could have introduced errors in the assignment of IC NDS to a proper river segment. To eliminate these type of errors, spatial results were manually verified for those sampling points where discrepancies between physical variables derived from IC dataset and derived from CCM2 dataset were unexpectedly large. Also sampling points located near river junctions were inspected. The manual inspection was done against the WISE WFD spatial dataset or, when this was not available, using Google Earth.

This verification meant that some sampling points previously assigned to a river were not assigned to any river segment anymore, if the river was not represented in the CCM2 database. These sampling points were excluded from further analysis. Also sampling sites from Germany, France and Romania were excluded from the CART analysis as no permission was granted to use the data.

It was observed that the location of rivers, and the boundaries of their catchments, were in some cases ill-represented (as compared to independent sources), which could generate some level of inaccuracy (**Figure 14**). These errors can affect river location, sinuosity and length, and also meant that sometimes catchments were incorrectly delimited who would have also affected catchment size. This was observed mainly in flat areas where borders between catchments and relief are less evident. However, we believe that errors resulting from using a common dataset are smaller than those from using Intercalibration attributes, which were collected manually by many observers. For the same reason it was decided to use Bio-geographical regions, instead of Intercalibration Regions.

Some of the geological formations, for example of 'Undifferentiated alluvial deposits (or glacial deposits)' type were ambiguous to classify either into siliceous or calcareous geological type. Some of those deposits were located in the vicinity of calcareous formations and the other in between siliceous formations. In this version of the map all of those deposits were classified as siliceous. This could be further differentiated if needed.

The non-disturbed sites (NDSs) were distributed unevenly across Europe, thus not all aspect of the abiotic variability had been represented in an equal manner. For example, from bio-geographic regions Pannonian Bio-geographical region included only one NDS, and others such as Anatolian, Arctic, Stepic, and Black Sea. Therefore, for representation of non-disturbed river habitat types the rivers located in those regions were excluded. The Pannonian Bio-geographic Region was included.

Another potential issue could be observed in the Continental Region dominated by FCMacHT 2. Here large rivers are not separated from the smaller rivers, as it is apparent for Boreal FCMacHTs. This can be a consequence of smaller differences in expected community compositions, insufficient data, or poor data quality from NDS on large rivers that correspond to the most densely populated areas of Europe. Large NDS sites are rarer and data collection from large, deep rivers is more complicated. This question can be only answered by more intense fisheries data collections on large rivers in this region.

The maps and benchmarks created here are an essential but intermediate step into assessment of barrier impact on aquatic habitats and fauna. It is expected that due to inherent biological variability it will be difficult to establish metrics for direct comparison of observed fauna with Expected Fish Communities. Such a comparison should be more robust at the habitat level, however the habitat characteristics used in this model are at a scale too coarse to be sensitive to small barrier impacts. Therefore, our next step will be to develop habitat model with mesoscale attributes (depth, velocity etc.) for the guilds presented here. The expected proportions of guilds for each FCMacHT class will allow us to calculate expected habitat proportions, which can then be more easily compared with habitat distribution upstream and downstream of barriers. Such habitat models will also permit us to incorporate temporal habitat fluctuations necessary for the analysis of climate change impacts.

5 REFERENCES

Bak, P. 1996. *How Nature Works.* Copernicus: New York, NY; 212.

Bain, M. B. & Meixler, M.S. 2008. A target fish community to guide river restoration. *River Research and Applications* 20:453–458.

Breiman, L., Freidman, J., Olshen, R. & Stone, C. 1984. *Classification and regression trees.* Wadsworth, Belmont, California, USA.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Ecology* 18:117–143.

De'ath, G. & Fabricius, K. E. 2002. Classification of regression trees: a powerful tool yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.

EFI+ Manual. 2009. *Manual for the application of the new European Fish Index—EFI+.* A fish-based method to assess the ecological status of European running waters in support of the Water Framework Directive. June 2009. EFI+ Consortium.

EEA .2012. European catchments and Rivers network system (Ecrins v1.1), European Environmental Agency, European Commission, Permalink: *https://www.eea.europa.eu/ds_resolveuid/401cd8dc-349e-4b38-8acc-d862460d0491*

EEA. 2017. WISE WFD reference spatial data set, Water Information System for Europe (WISE), Surface water body, European Environmental Agency, European Comission, Permalink: *https://www.eea.europa.eu/ds_resolveuid/eb5dc9ac13c342959d7733be79137fea*

EU Water Framework Directive. 2000. Directive of the European Parliament and the Council 2000/60/EC establishing a framework for community action in the field of water policy. *Official Journal of the European Communities 22.12.2000 L 327/1.*

European Environment Agency. 2016. Biogeographical regions for Europe 2016, version 1 (EEA, 2016). The European Soil Database distribution version 2.0, European Commission and the European Soil Bureau Network, CD-ROM, EUR 19945 EN, 2004, accessed from: https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871 .

Hastie, T., Tibshirani, R. & Friedman, J. 2001. *The Elements of Statistical Learning*, Springer, pp.468–469.

Holzer, S. 2008. European Fish Species: Taxa and Guilds Classification Regarding Fish-Based Assessment Methods. *Diplomarbeit zur Erlangung des akademischen Grades Magistra der Naturwissenschaften (Mag.rer.nat.) an der Fakultät für Lebenswissenschaften der Universität Wien Wien*, im November 2008.

Kaufman, L. & Rousseeuw, P. J. 1987. *Clustering by means of Medoids, in Statistical Data Analysis Based on the L_1–Norm and Related Methods*, edited by Y. Dodge, North-Holland, 405–416.

Krause, E. F. 1987. *Taxicab Geometry*. Dover. ISBN 0-486-25202-7.

Kuehne, R. A. 1962. A classification of streams, illustrated by fish distribution in an eastern Kentucky creek. *Ecology* 43:608-614.

Lyons, J., Stewart, J. S. & Mitro, M. 2010. Predicted effects of climate warming on the distribution of 50 stream fishes in Wisconsin, U.S.A. *Journal of Fish Biology* 77:1867–1898

Logez, M., Bady, P., Melcher, A. & Pont, D. 2012. A continental-scale analysis of fish assemblage functional structure in European rivers. *Ecography* 35:001–012.

Matthews, W. J. & Marsh-Matthews, E. 2017. *Stream Fish Community Dynamics: A Critical Synthesis.* JHU Press. pp 345.

Lyons, J., Stewart, J. S. & Mitro, M. 2010. Predicted effects of climate warming on the distribution of 50 stream fishes in Wisconsin, U.S.A. *Journal of Fish Biology* 77:1867–1898.

McKenna, J. E. Jr, McDonald, R. P., Castiglione, C., Morrison, S., Kowalski, K. & Passino-Reader, D. 2006. A broadscale fish–habitat model development process: Genesee Basin, New York. Pages 533–554 *In* Hughes, R. M., Wang, L. & Seelbach, P. W. (eds) *Landscape influences on stream habitats and biological assemblages.* American Fisheries Society, Symposium 48, Bethesda, Maryland.

McKenna, J. E. Jr, Schaeffer, J. S., Stewart, J. E. & Slattery, M. T. 2015. Development of a spatially universal framework for classifying stream assemblages with application to conservation planning for Great Lakes lotic fish communities. *Restoration Ecology* 23:167–178.

Melcher, A., Schmutz, S., Haidvogl, G. & Moder, K. 2007. Spatially based methods to assess the ecological status of European fish assemblage types. *Fisheries Management and Ecology* 14:453–463.

Omernik, J.M. & Griffith, G.E. 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environmental Management* 54:1249-1266.

Panagos, P. 2006. The European soil database (2006) *GEO: connexion* 5(7):32-33.

Parasiewicz, P., Pegg, M., Rogers, J., Behmer, A. & Eldridge, A. 2012. Developing environmental flows for fish and wildlife: a mesohabitat study on the Niobrara River. Project completion report to Nebraska Game and Parks Commission, Lincon, NE.

Park, Y. S., Oberdorff, T. & Lek, S. 2005a. Patterning riverine fish assemblages using an unsupervised neural network. *In* Lek, S., Scardi, M., Verdonschot, P. F. M., Descy, J-P. & Park, Y. S., *Modelling Community Structure in Freshwater Ecosystems.*

Park, Y. S., Verdonschot, P. F. M. & Lek, S. 2005b. Review of modelling techniques. *In* Lek, S., Scardi, M., Verdonschot, P. F. M., Descy, J-P. & Park, Y. S., *Modelling Community Structure in Freshwater Ecosystems.*

Park, H. S. & Jun, C. H. 2009. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications* 36:3336–3341.

Prus, P., Wiśniewolski, W. & Adamczyk, M. (*eds)*. 2016. *Monitoring ichtiofauny w rzekach*. *Przewodnik metodyczny.* [*Ichthyofauna monitoring in rivers. A methodological guide.] Biblioteka Monitoringu Środowiska.* Główny Inspektorat Ochrony Środowiska, Warszawa, Poland, 95 pp.

Polish National Water Management Authority. 2015. Ustalenie metody szacowania przepływów środowiskowych w Polsce. *Etap II Raport końcowy*. PS-00-R-010-02. Warsaw, 186 pp.

Qian, S. S. 2016. *Environmental and Ecological Statistics with R.* CRC Press, 536 pp, second edition.

Steen, P.J., Zorn, T.G., Seelbach, P.W. & Schaeffer, J.S. 2008. Classification tree models or predicting distributions of Michigan stream fish from landscape variables. *Transactions of the American Fisheries Society* 137:976–996

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union* 38:913-920.

WFD Intercalibration. 2011. Phase 2: *Milestone report – October 2011.* European Commission Directorate General, JRC Joint Research Centre, Institute of Environment and Sustainability. JRC documents, 105 pp.

Welcomme, R. L., Winemiller, K. O. & Cowx, I. G. 2006. Fish environmental guilds as a tool for assessment of ecological condition of rivers. *River Research and Application* 22:377-396.

Vogt, J. V. , Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M-L., Haastrup, P. & Bamps, C. 2007. A pan-European River and Catchment Database. *European Commission - JRC*, Luxembourg, (EUR 22920 EN), 120 pp.