

1 Title

2 Disentangling the effects of land use and geo-climatic factors on diversity in European
3 freshwater ecosystems

4

5 Authors

6 Christian K. Feld ^{1‡}, Sebastian Birk ¹, David Eme ², Michael Gerisch ³, Daniel Hering ¹,
7 Martin Kernan ⁴, Kairi Maileht ⁵, Ute Mischke ⁶, Ingmar Ott ⁵, Florian Pletterbauer ⁷, Sandra
8 Poikane ⁸, Jorge Salgado ⁴, Carl D. Sayer ⁴, Jeroen van Wichelen ⁹ and Florian Malard ²

9

10 Affiliations

11 ¹ Faculty of Biology, Department of Aquatic Ecology and Centre for Water and
12 Environmental Research (ZWU), University of Duisburg-Essen, Universitätsstr. 5, 45141
13 Essen, Germany

14 ² Université de Lyon ; UMR5023 Ecologie des hydrosystèmes Naturels et Anthropisés ;
15 Université Lyon1 ; ENTPE ; CNRS ; 6 rue Raphaël Dubois, 69622 Villeurbanne, France

16 ³ Department of Conservation Biology, Helmholtz Centre for Environmental Research – UFZ,
17 Permoserstr. 15, 04318 Leipzig, Germany

18 ⁴ Environmental Change Research Centre, Department of Geography, University College
19 London, Gower Street, London WC1E 6BT, UK

20 ⁵ Institute of Agricultural and Environmental Sciences, Centre for Limnology, Estonian
21 University of Life Sciences, Rannu Parish, Tartu 61117 Estonia

22 ⁶ Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries
23 (IGB), Müggelseedamm 310, Berlin 12587 Germany

24 ⁷ Institute of Hydrobiology and Aquatic Ecosystem Management, University of Natural
25 Resources and Life Sciences Vienna, Max Emanuel Straße 17, 1180 Vienna, Austria

26 ⁸ European Commission, Joint Research Centre, Institute for Environment and Sustainability,
27 via E. Fermi 2749, Ispra, VA I-21027 Italy

28 ⁹ Protistology and Aquatic Ecology, Ghent University, Krijgslaan 281 – S8, Gent 9000,
29 Belgium

30

31 Highlights

- 32 • diversity in eleven organism groups across five aquatic ecosystems was quantified
- 33 • land use alone explained little variation in aquatic biodiversity
- 34 • geo-climatic (natural) descriptors explained significantly more variation
- 35 • land use interacted strongly with natural descriptors
- 36 • natural descriptor's effects were not driven by spatial gradients

37

38 Corresponding author:

39 Christian K. Feld, Department of Aquatic Ecology, Faculty of Biology and Centre for Water
40 and Environmental Research (ZWU), Universitätsstr. 5, 45141 Essen, Germany; E-mail:

41 christian.feld@uni-due.de, phone: +49 (0) 201 183-4390

42 **Abstract**

43 Land use effects are considered among the main stressors on freshwater biodiversity, with up
44 to 80% of land in Europe under intensive use. Here, we address the impact of arable and
45 urban landscapes on taxon richness, Shannon-Wiener diversity, taxon rareness and taxonomic
46 distinctness of eleven organism groups encompassing vertebrates, invertebrates and plants,
47 occurring in five freshwater ecosystem types across Europe: rivers, floodplains, lakes, ponds
48 and groundwater. In addition, nine geo-climatic descriptors (e.g. latitude, longitude,
49 precipitation) were used to disentangle land use effects from those of natural drivers of
50 biodiversity. Using a variance partitioning scheme based on boosted regression trees and
51 generalised linear regression modelling, we sought: i) to partition the unique, shared and
52 unexplained variation in the metrics explained by both groups of descriptor variables, ii) to
53 quantify the contribution of each descriptor variable to biodiversity variation in the most
54 parsimonious regression model and iii) to identify interactions of land use and natural
55 descriptors. The variation in biodiversity uniquely described by land use was consistently low
56 across both ecosystem types and organism groups. In contrast, geo-climatic descriptors
57 uniquely, and jointly with land use, explained significantly more variance in all 39
58 biodiversity metrics tested. Regression models revealed significant interactions between geo-
59 climatic descriptors and land use for a third of the models, with interactions accounting for up
60 to 17% of the model's deviance. However, no consistent patterns were observed related to the
61 type of biodiversity metric and organism group considered. Subdividing data according to the
62 strongest geo-climatic gradient in each dataset aimed to reduce the strength of natural
63 descriptors relative to land use. Although data sub-setting can highlight land use effects on
64 freshwater biodiversity, sub-setting our data often failed to produce stronger land use effects.
65 There was no increase in spatial congruence in the subsets, suggesting that the observed land
66 use effects were not dependent upon the spatial extent of the subsets. Our results confirm
67 significant joint effects of, and interactions between, land use and natural environmental

68 descriptors on freshwater biodiversity, across ecosystem types and organism groups. This has
69 implications for biodiversity monitoring. First, the combined analysis of anthropogenic and
70 natural descriptors is a prerequisite for the analysis of human threats to biodiversity. Second,
71 geo-climatically, but not necessarily geographically more homogeneous datasets can help
72 unmask the role of anthropogenic descriptors. And third, whole community-based
73 biodiversity metrics (including taxon richness) are not ideal indicators of anthropogenic
74 effects on biodiversity at broad scales.

75

76 Keywords

77 Freshwater ecosystems, biodiversity, arable land, urban area, geo-climatic descriptors,
78 variance partitioning

79 **1. Introduction**

80 Although freshwaters cover only 1% of the earth's surface, almost 10% of the world's species
81 live in freshwater ecosystems (Loh & Wackernagel, 2004). Freshwater biodiversity is
82 declining faster than marine and terrestrial biodiversity (Dudgeon *et al.*, 2006), most likely
83 because human life and many human activities rely on fresh water. This results in high
84 population densities, intense land and water uses and modification and pollution hotspots in
85 the vicinity of freshwater bodies. Consequently, human impacts on freshwater biodiversity are
86 numerous and wide-ranging. Dudgeon *et al.* (2006) identify five major stressors of
87 biodiversity which affect different freshwater ecosystem types to varying degrees: i) water
88 overexploitation; ii) water pollution; iii) flow modification; iv) habitat degradation; and v)
89 invasive species. While rivers are more affected by physical alterations (e.g. dams,
90 impoundments, disconnection from the floodplain), lentic waters are more susceptible to
91 nutrient enrichment (Wetzel *et al.*, 2001; Schindler, 2006), with increasing adverse effects on
92 lentic biota under climate change (Jeppesen *et al.*, 2010; 2012).

93 Numerous stressors are linked to land use, which therefore is considered a composite (or
94 proxy) stressor. Intensive agriculture, in particular, affects both lotic and lentic biodiversity
95 through flow modification, pollution by fine sediment and pesticide fluxes (Allan, 2004; Feld,
96 2013), habitat degradation and eutrophication (Jeppesen *et al.*, 2000). Urbanisation represents
97 another intensive land use, with strong effects on freshwater biodiversity, resulting in
98 "consistent declines in the richness of algal, invertebrate, and fish communities" (Paul &
99 Meyer, 2001). In Europe, a very high proportion (up to 80%) of the land is intensively used
100 for settlements, infrastructure and production systems (including agriculture and intense
101 forestry: <http://www.eea.europa.eu/themes/landuse/intro>; accessed on 11 June 2015) and
102 aquatic biodiversity is probably impoverished accordingly. Because of this cocktail of
103 stressors, freshwater ecosystems and their biodiversity are currently among the most

104 threatened on the planet, prompting scientists and politicians to develop strategies to **sustain**
105 **and improve** biodiversity functioning and ecosystem service provisioning.

106 Anthropogenic stress intensity and thus its influence on biodiversity differs regionally,
107 impacting large-scale biodiversity patterns, originally shaped by natural drivers. These natural
108 drivers are considered in macro-ecological and other broad-scale studies highlighting the role
109 of i) energy/climate (e.g. Mittelbach *et al.*, 2007; Pearson & Boyero, 2009; Heino, 2011), ii)
110 area/habitat heterogeneity (e.g. Vinson & Hawkins, 1998; Oberdorff *et al.*, 2011) and iii)
111 history (e.g. Leprieur *et al.*, 2011; Vinson & Hawkins, 2003). The influence that energy and
112 climate have on biodiversity are primarily driven by temperature, precipitation and evapo-
113 transpiration, all of which influence ecosystem energy supply and thus control or support
114 biophysical processes operating within the system (Wright, 1983; Hawkins *et al.*, 2003; Evans
115 *et al.*, 2005; Mittelbach *et al.*, 2007; Field *et al.*, 2009). However, temperature and evapo-
116 transpiration vary with altitude, and more importantly, freshwater biodiversity is also found to
117 increase with altitude (see Vinson and Hawkins, 1998 for a review on benthic invertebrates).
118 This suggests temperature is unlikely to be the main co-variate of the energy/climate driver in
119 freshwater ecosystems, and in more general terms, the role of energy/climate differs between
120 terrestrial and aquatic systems (Field *et al.*, 2009).

121 Area/habitat heterogeneity refers to the size and heterogeneity (habitat diversity) of an area
122 under consideration, with the assumption that larger and more heterogeneous areas exhibit
123 higher biodiversity (*sensu* MacArthur & Wilson, 1963; Guégan *et al.*, 1998; Davies *et al.*,
124 2007). Lastly, historical events (i.e. previous and often long-term events dating back for
125 centuries or even millennia) may continue to shape contemporary biodiversity patterns
126 (Mittelbach *et al.*, 2007; Leprieur *et al.*, 2011; Tisseul *et al.*, 2012). The expansion of
127 Pleistocene glaciers and their subsequent contraction followed by recolonisation, for example,
128 are considered a key factor in explaining much of the variation in the distribution of

129 contemporary biodiversity across Europe (Reyjol *et al.*, 2007; Araujo *et al.*, 2008; Baselga *et*
130 *al.*, 2012), with formerly glaciated regions (e.g. Scandinavia) generally exhibiting less
131 diversity than non-glaciated regions (e.g. Mediterranean peninsula). Over more recent
132 timescales land use practices dating back decades may continue to shape contemporary
133 biodiversity even if land use has subsequently changed or been abandoned (Harding *et al.*,
134 1998).

135 Both the natural drivers of freshwater biodiversity and multiple stressors resulting from
136 human land and water uses have been addressed in many studies (see Stendera *et al.*, 2012 for
137 a recent summary of 368 papers), although few have considered these in an integrated way.
138 Studies that investigate the combined effects of natural and anthropogenic descriptors are
139 rare, but are necessary to address metacommunity aspects in ecosystem assessment studies
140 (Heino 2013). Furthermore, Stendera *et al.* (2012) found that the majority of studies on
141 natural drivers were rather broad-scale (continental and global), whereas studies on
142 anthropogenic stressors tend to focus on much finer (regional and local) spatial scales. The
143 spatial resolution (grain size) also often differs, with the catchment ‘grain’ prominent in
144 broad-scale studies, but single sites within one or several catchments foremost in fine-scale
145 studies. The mechanisms driving biodiversity, however, are likely to vary with spatial grain
146 (local ecosystem vs. catchment) and extent (Field *et al.* 2009, Heino 2011). Few studies
147 addressed the impacts of both natural drivers and anthropogenic stressors on freshwater
148 biodiversity (Irz *et al.*, 2007; Argillier *et al.*, 2013; Brucet *et al.*, 2013) and there remains a
149 limited understanding of the synergies between both groups of descriptors.

150 In this study, we developed a stepwise analysis to determine the independent, overlapping and
151 interacting effects of land use and geo-climatic variables (hereafter referred to as descriptors)
152 on the European biodiversity patterns of eleven organism groups in five lotic and lentic
153 ecosystem types (rivers, lakes, floodplains, ponds and groundwater). We used a machine-

154 learning technique to partition the variance and to quantify the independent and overlapping
155 effects of both descriptor groups in each ecosystem. In line with previous studies at
156 continental scale (e.g. Bruce *et al.*, 2013), we hypothesised a strong influence of natural
157 descriptors on biodiversity (e.g. latitude, mean annual temperature), but a much weaker role
158 of agricultural and urban land uses. As land use, however, is not independent of, for example,
159 altitude (i.e. slope), temperature and precipitation, we expected strong joint effects. This was
160 analysed by variance partitioning, and further tested by means of significant interaction terms
161 between single land use and geo-climatic descriptor variables in regression modelling. To
162 decrease the effect of the most influential geo-climatic descriptor in the regression models, we
163 generated subsets of the data and quantified the proportion of variance attributable to land use
164 separately for each subset. This procedure was driven by the hypothesis that geo-climatically
165 more homogeneous data (with shorter natural gradients) would reveal a stronger influence of
166 land use on biodiversity. In order to account for the response of different aspects of
167 biodiversity, we compared the results of four widely used biodiversity metrics: taxon
168 richness, Shannon-Wiener diversity, taxon rareness and taxonomic distinctness (Clarke and
169 Warwick 1998). The first two metrics quantify the number and equal distribution of species
170 within a community and thus represent very basic concepts of diversity, i.e. richness and
171 equity. The latter two metrics add the aspects of relative rareness of taxa and their
172 phylogenetic relationships to each other within a community. We hypothesised that taxa are
173 not equally sensitive to human impact and that in particular those taxa become rare in stressed
174 systems that respond already to low stress levels, i.e. sensitive taxa. Sensitive and insensitive
175 (tolerant) taxa, however, are not equally distributed among taxonomic families, orders and
176 higher taxonomic units. Following Warwick and Clarke (1995), we thus anticipated stressed
177 systems to remain in an early successional stage, with guilds of closely related insensitive
178 taxa and thus with a low taxonomic distinctness compared to unstressed systems.

179 This is the first study to quantify and disentangle the response of different biodiversity aspects
180 to natural and human-induced environmental descriptors, across five lotic and lentic
181 freshwater ecosystems and eleven organism groups, and applying the same analytical
182 approach.

183 **2. Material and Methods**

184 *2.1 Anthropogenic descriptors*

185 For all but groundwater ecosystems we used CORINE land cover data (European
186 Environmental Agency; <http://www.eea.europa.eu/publications/COR0-landcover>) to calculate
187 the proportion of arable and urbanised land within a catchment or the area directly
188 surrounding a site (Table 1). The area considered differed between ecosystem types and was
189 selected to match the scale of biological sampling. The CORINE land cover data are based on
190 satellite imagery (Landsat 7, 25 x 25 m pixels), cover most countries in Europe (geometric
191 accuracy: 100 m) and encompass land cover types with a minimum area of 25 ha. We used
192 the land cover classes 'arable land' and 'urban land' (hereafter referred to as land use), which
193 aggregate the CORINE level 3 types '2.1.1 Non-irrigated arable land' as 'arable land' and the
194 level 2 types '1.1 Urban fabric' and '1.2 Industrial, commercial and transport units' as 'urban
195 land'. We focused on these two land use types, because they are known to strongly affect
196 aquatic biodiversity via numerous individual stressors (Paul & Meyer, 2001; Allan, 2004;
197 Feld, 2013). For groundwater systems, we used the GlobCover land cover data
198 (<http://due.esrin.esa.int/globcover/>) due to its comprehensive coverage of Eastern Europe.
199 'GlobCover Land Cover v2' is a global land cover map at a resolution of 10 arc seconds (or
200 300 m at the equator) and corresponds well with the CORINE land cover classification.
201 Arable and urban land uses were projected onto an ETRS-Lambert Azimuthal Equal Area
202 grid covering Europe with a cell (grid) size of 100 x 100 km (EDIT geoplatform). The same
203 grid was applied to generate the land use data for lakes using the CORINE land cover data.

204 Proportions of different land use types were obtained by clipping the land use maps (either
205 CORINE or GlobCover) with a layer containing the polygonal information from the targeted
206 areas (Table 1) within a geographic information system (ESRI ArcGIS 10, Redlands, CA).

207 *2.2 Geo-climatic descriptors*

208 We used nine natural environmental descriptor variables covering geographical and climatic
209 aspects (hereafter summarised as geo-climatic descriptors, Supplementary Table S1). We did
210 not separate geographical and climatic variables in our analysis as the objective was to
211 compare the strength of natural vs. anthropogenic descriptors. Furthermore, geographical and
212 climatic variables tend to be strongly correlated at the spatial scale of the data employed in
213 this study (e.g. mean annual air temperature, latitude and altitude). Latitude, longitude,
214 altitude and catchment size were derived from digital maps using ArcGIS 10. Latitude and
215 longitude were included as proxy geographical variables representing other potential natural
216 drivers of biodiversity, such as historical climate and glaciation (Hortal *et al.*, 2011; Stendera
217 *et al.*, 2012), but were excluded from the analysis if they were collinear with any of the other
218 environmental descriptors (compare Supplementary Table S1). Altitude was included to
219 account for the role of topography in shaping diversity patterns (e.g. Davies *et al.*, 2006).
220 Lake surface area was derived from the WISER lake database (Moe *et al.*, 2013). Mean
221 annual air temperature and annual precipitation were abstracted from the WorldClim database
222 version 1.4 (Hijmans *et al.*, 2005). WorldClim summarises measured data at weather stations
223 between 1950 and 2000 as monthly mean values, interpolated by a thin-plate smoothing
224 spline algorithm to fit a raster grid (grid size: 30 arc seconds, approximately 1 km at the
225 equator). Mean annual air temperature was averaged from long-term yearly means, whereas a
226 yearly mean was averaged from monthly means throughout a year. Annual precipitation was
227 based on the sum of long-term monthly mean precipitation values. Actual and potential

228 evapo-transpiration (AET, PET) were derived from the CGIAR-CSI Global-PET database
229 (for details, see Zomer *et al.*, 2008; <http://www.cgiar-csi.org>).

230 *2.3 Biodiversity data*

231 2.3.1 Rivers

232 Site-specific river data were derived from the WISER river database (Moe *et al.*, 2013),
233 encompassing taxa lists of fish, macroinvertebrate and macrophyte communities and
234 proportional catchment land use for up to 1,221 sites across Central Europe (Central/Western
235 Mountains and Central/Western Plains ecoregions of France, Germany and Austria; Illies,
236 1978). Macroinvertebrate data were available for all sites, fish data for 590 sites and
237 macrophyte data for 651 sites. The taxa lists originate from national monitoring surveys and
238 followed the national monitoring standards defined for field sampling methodology and
239 sample processing (see Dahm *et al.*, 2012 and Feld, 2013 for details). If multiple samples
240 were available for the same site, we used only the most recent data, with spring samples
241 preferred (March–May) as this is the most appropriate season for sampling small and mid-
242 sized wadable streams <500 km² catchment area (>85% of all river sites).

243 Prior to the calculation of biodiversity metrics, the raw taxa lists obtained from the WISER
244 river database were manually adjusted to eliminate researcher-dependent bias, for example,
245 caused by different taxonomic determination levels for macroinvertebrates (e.g. Oligochaeta,
246 Diptera). Species-level identification was achieved for fish and macrophytes, while genus
247 level was used for macroinvertebrates, as this is the standard determination level in France
248 (see Table 1 for the overall taxon richness in the river data).

249 2.3.2 Lakes

250 Lake-specific phytoplankton taxa lists from 836 lakes (surface area >0.5 km², reservoirs
251 excluded) in 20 European countries were derived from the WISER lake database (Moe *et al.*,
252 2013). The lakes are distributed among three major European regions: i) the Mediterranean

253 region (145 lakes in Cyprus, Italy, Spain, Portugal and Romania), ii) the Central/Baltic region
254 (373 lakes in Belgium, Germany, Estonia, France, Hungary, Lithuania, Latvia, the
255 Netherlands and Poland) and iii) the Northern region (318 lakes in Denmark, Finland, Ireland,
256 Norway, Sweden and the United Kingdom).

257 Lake taxa lists (overall taxon richness given in Table 1) originate from national monitoring
258 data and international research projects (Moe *et al.*, 2013). We used samples taken between
259 2004 and 2010 to maximise the temporal comparability of samples and selected only the most
260 recent data within this period, if multiple year samples were available for a lake. Further, we
261 used a single lake sample encompassing data from all sampling points within the same water
262 body and/or different sampling periods within a year, which were combined by averaging to
263 create a mean abundance for each lake. Prior to the calculation of biodiversity metrics, taxa
264 records from each country were harmonised for nomenclature (Phillips *et al.*, 2013).

265 2.3.3 Ponds

266 We defined ponds as shallow lentic water bodies with surface area less than five hectares
267 (0.05 km²) (De Meester *et al.*, 2005). Pond taxa lists were obtained from 32 peer-reviewed
268 publications indexed in the Web of Science and generated for amphibians, macroinvertebrates
269 (Gastropoda, Odonata and Coleoptera only) and macrophytes (for overall taxon richness see
270 Table 1). Data were also collated from Homes, Hering & Reich (1999), Nagorskaya *et al.*
271 (2002), Sobkowiak (2003), Oertli *et al.* (2005), Sayer *et al.* (2012), Böhmer (2012), Moe *et al.*
272 (2013), the European Pond Conservation Network (University of Sevilla, ES; unpubl.), N.J.
273 Willby (University of Stirling, UK; unpubl.) and B.A. Lukács (Hungarian Academy of
274 Sciences, HU; unpubl.). All data were sampled focussing on representative biological
275 inventories within the scope of biodiversity-studies (e.g. Oertli *et al.*, 2005). Amphibia were
276 generally surveyed in spring to obtain a full list of species occurring in the pond. Gastropoda
277 and Coleoptera were sampled during summer, employing stratified sampling of all dominant

278 habitats using a hand-net. Adult stages were recorded for Odonates, usually merging the
279 sampling efforts of at least two occasions in early and late summer. Macrophyte records were
280 generally obtained during the summer season from transects covering representative pond-
281 habitats.

282 Amphibian species data were recorded at 148 ponds in seven European countries (Belgium,
283 France, Germany, Italy, Poland, Spain, Switzerland). Macroinvertebrate taxa lists were
284 collated using 189 samples from 176 ponds in eleven countries (Czech Republic, Estonia,
285 France, Germany, Ireland, Italy, Poland, Spain, Sweden, Switzerland, United Kingdom).
286 Macrophyte species records comprised 601 samples at 392 ponds in seven countries
287 (Belgium, Germany, Hungary, Poland, Sweden, Switzerland, United Kingdom). We
288 considered only species classified as ‘hygrophytes’, ‘helophytes’ and ‘hydrophytes’ with
289 Ellenberg’s moisture values ≥ 7 and stoneworts (Ellenberg *et al.*, 1992).

290 Species level was achieved for amphibia, the majority of macroinvertebrates and most
291 macrophyte taxa (except for *Chara* sp. and *Callitriche* sp.). We generated presence/absence
292 data only from the pond taxa lists, because abundance data were not available in all studies.

293 2.3.4 Floodplains

294 The floodplain taxa lists were derived from the peer-reviewed literature of European datasets
295 on ground beetles, molluscs and higher plants (floodplain vegetation) in riverine wetland
296 ecosystems (overall taxon richness given in Table 1). We reviewed relevant publications
297 between 1990 and 2012 using the Web of Science. This resulted in 78 publications useful to
298 generate the three taxa lists (total number of sample sites: 565): 132 sites for ground beetles,
299 81 sites for molluscs and 352 sites for floodplain vegetation. The sites are located in 21
300 countries and on 51 river floodplains across Europe, with the majority of sites located in
301 Central Europe: Poland: 99 sites, Germany: 98, France: 81, Belgium: 42, Switzerland: 29, the
302 Netherlands: 25, Czech Republic: 7 and Denmark: 6.

303 Standardisation of species abundances among studies was impossible due to the use of
304 different sampling methods. To minimise the bias in sampling effort, we omitted studies with
305 an extremely short or long field sampling period and those with strongly skewed or otherwise
306 inconsistent data.

307 2.3.5 Groundwater

308 Site-specific groundwater data were derived from the European groundwater crustacean
309 dataset recently published by Zigmajster *et al.* (2014). It covers the whole of Europe, except
310 Russia, and contains a total of 21,700 occurrence data, which collectively represent 12 orders
311 and 1,570 species and subspecies of obligate groundwater Crustacea. Occurrence data are
312 restricted to species that complete their entire life cycle exclusively in groundwater, but they
313 are from a variety of habitats in karst and porous aquifers and the hyporheic zone of streams.
314 They were projected in ETRS -Lambert Equal Area onto a grid system containing 494 cells of
315 100 x 100 km. This spatial resolution ensured that the species richness pattern of groundwater
316 crustaceans was not biased by variable sampling effort among cells (see Zigmajster *et al.*,
317 2014).

318 *2.4 Calculation of biodiversity metrics*

319 Biodiversity has many facets and, amongst others, encompasses compositional (structural),
320 functional (trait) and phylogenetic aspects of assemblages. Given the mixture of binary
321 (presence/absence) and continuous (abundance) data, we restricted the set of biodiversity
322 metrics used here to total species richness, species rareness and taxonomic distinctness (i.e. a
323 proxy for phylogenetic diversity). With abundance data, we also calculated Shannon-Wiener
324 diversity (hereafter referred to as Shannon diversity). Species richness and Shannon diversity
325 are among the most commonly-used indicators of aquatic biodiversity in Europe (see Birk *et*
326 *al.*, 2012 for a recent review of monitoring methodology) and basically account for two
327 structural characteristics of a community: the number and the equal distribution of its

328 members. Low diversity may be linked to environmental impacts, for example, intensive land
329 use or other sources of habitat deterioration, which can act as landscape filters on species and
330 communities (Poff, 1997). Yet, the general utility of such community-based diversity metrics
331 as indicators of environmental deterioration has been subject to criticism, because it neglects
332 species identity. Species composition may change along habitat deterioration gradients, with
333 largely constant values for richness and Shannon diversity (Feld *et al.*, 2013). We, therefore,
334 introduced taxon rareness (similar to the index of endemism proposed by Crisp *et al.*, 2001
335 and Linder, 2001), which describes the sum of relative frequencies of all taxa encountered at a
336 site or within an area (grid) in relation to the overall number of sites or areas (grids) where the
337 individual taxa have been observed. Hence, the index provides a measure of the summed
338 relative frequencies of ‘endemic’ (or rare) taxa within a community, based on the overall
339 frequency of the taxa in the entire dataset. Taxon rareness does not account for taxon identity
340 either, but weights rare taxa, which are believed to respond very sensitively to any kind of
341 habitat deterioration and hence are among the first to disappear under enhanced
342 environmental stress. We included taxonomic distinctness in this study to address the
343 phylogenetic connections of the taxa within a community. Community members representing
344 many different families, orders and classes along a Linnéan phylogenetic tree are
345 phylogenetically less related to each other, i.e. they are taxonomically distinct. For example,
346 three species of the same genus are taxonomically less distinct than three species of different
347 genera, orders or higher taxonomic entities, which is why taxonomic distinctness is also
348 referred to as phylogenetic diversity. The index of taxonomic distinctness proposed by Clarke
349 & Warwick (1998, 1999) calculates the mean taxonomic dissimilarity of any pair of taxa
350 within a community along the Linnéan phylogenetic tree (i.e. species, genus, family, order,
351 class and phylum). The index is applicable to binary taxa lists and adds a unique aspect of
352 biodiversity, covered neither by taxon richness nor by taxon evenness (Heino *et al.* 2005;
353 Gallardo *et al.*, 2011; Feld *et al.*, 2013).

354 *2.5 Data analysis*

355 We applied a stepwise analytical protocol for the multivariate analysis using Boosted
356 Regression Tree analysis (BRT) and Generalised Linear Modelling (GLM). Both methods
357 complement each other and address specific aspects of the analytical approach. BRT was used
358 to partition the variation in the biodiversity metrics explained by geo-climatic and land use
359 descriptors as it is capable of handling collinear data of different numerical scales in the same
360 analysis. It was thus possible to undertake comparable variation partitioning for each of the
361 combination of eleven organism groups and up to four diversity metrics. Following the results
362 of BRT (see below), GLM was used to identify the most parsimonious model for each
363 biodiversity metric, i.e. the model with the least number of significant predictors, including
364 significant interaction terms. Hence, GLM allowed us to analyse the role and significance of
365 each descriptor, and its interaction with others, in more detail.

366 *2.5.1 STEP 1*

367 Individual BRTs were run for each possible combination of organism group and biodiversity
368 metric using all geo-climatic and land use descriptors (full model) to compare the effects of
369 both descriptor groups. BRT constitutes a machine-learning method that combines classical
370 regression (decision) tree analysis with boosting (Elith, Leathwick & Hastie, 2008). Decision
371 trees are intuitive and easy to visualise, but very sensitive to changes in the environmental
372 descriptor variables. Because of the hierarchical structure of descriptors, any change in a
373 higher-level descriptor implies changes to connected descriptors located at lower hierarchical
374 levels in a decision tree. With BRT, boosting aims to improve the predictive performance of a
375 regression tree model by adding subsequent regression trees to the residuals of a former
376 model. Hence, it is possible to identify the model with the maximum deviation explained by
377 the descriptor variables.

378 The major advantages of BRT over classical regression modelling are its capacity to i)
379 analyse binary, ordinal and continuous descriptor variables, ii) accommodate collinear data,
380 iii) handle non-linear descriptors with missing values and iv) identify interactions between
381 descriptors (Elith *et al.*, 2008).

382 The full BRT models allowed us to identify the contribution of each individual descriptor to
383 the overall variance explained in a biodiversity metric as well as the pairwise interactions
384 between descriptors. Partial Dependence Plots (PDP) enabled the identification of the
385 response patterns of biodiversity metrics along environmental descriptor gradients (Cutler *et*
386 *al.*, 2007). PDPs helped identify potential thresholds along the geo-climatic gradients at which
387 a biodiversity metric value either sharply increased or decreased (Clapcott *et al.*, 2012; Feld,
388 2013a). Such thresholds may mark natural split points in the data, for example, geographical
389 splits at a specific latitude, longitude or altitude, which then imply the presence of spatial
390 patterns in the targeted biodiversity metric. We subsequently used these split points for the
391 strongest geo-climatic descriptor in each BRT to divide each dataset (i.e. ecosystem
392 type/organism group/biodiversity metric) into two subsets (step 3).

393 In addition to the full BRT models, we applied an additive partial regression scheme
394 (Legendre & Legendre, 1998, p. 531) to decompose the explained variation of the biodiversity
395 metrics into four fractions: i) pure geo-climatic, ii) pure land use, iii) shared geo-climatic/land
396 use and iv) unexplained. The shared fraction (iii) represents the variation that may be
397 attributed to geo-climatic and land use descriptors together and is obtained additively in
398 partial regression. As such, it is inherently different from interaction terms (multiplicative
399 relations) as introduced into the GLM (step 2). Differences in the variance explained were
400 tested for significance using a Wilcoxon signed rank test for paired samples.

401 2.5.2 STEP 2

402 GLM was applied individually to each combination of eleven organism groups and three
403 biodiversity metrics (Shannon diversity excluded), and a set of geo-climatic and land use
404 descriptors that excluded highly collinear variables, defined as those with a variance inflation
405 factor >8 (Zuur, Ieno & Smith, 2007). We choose GLM for this step because of its flexibility
406 in identifying the most parsimonious model (i.e. the best trade-off between model fit and
407 complexity), including interactions between anthropogenic and geo-climatic descriptors.
408 Adjusted goodness of fit (R^2) and Akaike Information Criterion (AIC) were used as GLM
409 quality indicators. The order of entry of each descriptor variable into a GLM model was based
410 on the individual explanatory strength of the variable as identified in step 1 using BRT (i.e.
411 the strongest descriptor was selected first, followed by the second strongest, and so on). This
412 procedure ensured a standardised and hence comparable analytical procedure for GLM
413 models for all ecosystems.

414 We used Poisson regression for species richness and Gaussian regression for
415 rareness/endemicity and taxonomic distinctness. If overdispersion was detected in Poisson
416 regression, we used negative binomial distribution functions in GLM. Rareness and
417 taxonomic distinctness were logit-transformed to better-fit Gaussian regression (Warton &
418 Hui, 2011). The GLM model with the highest explained adjusted deviance in combination
419 with the lowest AIC obtained for each combination of organism group and biodiversity
420 metric, was selected as the final model. A final model included marginally significant
421 descriptors ($0.05 < P < 0.1$), if the explained deviance and/or AIC notably improved with the
422 descriptors in the model.

423 2.5.3 STEP 3

424 The final analytical step repeated the procedure for step 2, but was applied to the data subsets.
425 These subsets were defined using the split points of the most influential geo-climatic
426 descriptor in each analysis. This was derived individually for each metric from the partial

427 dependence plots of the BRTs (step 1). If necessary, the split points were slightly adjusted, to
428 better achieve a balanced sample size from both data subsets. The objective of splitting the
429 data according to the most influential geo-climatic descriptor was to control for the variance
430 driven by the respective geo-climatic descriptor and thus to focus more on the role of land
431 use.

432 All statistical analyses were run in R 2.15.3 (R Development Core Team, 2013). For BRTs,
433 we used the packages ‘gbm’ (Ridgeway, 2013) and ‘dismo’ (Hijmans *et al.*, 2013). GLMs
434 were run with the package ‘MASS’ (Venables & Ripley, 2002).

435 **3. Results**

436 *3.1 Comparison of geo-climatic descriptors and land use*

437 Across all ecosystems, eleven organism groups were analysed resulting in 39 separate
438 biodiversity responses (Figure 1). Together, geo-climatic and land use descriptors explained
439 between 20 and 93% (mean: 35%, SD: 18.7%) of the total variance in the full BRT models.
440 On average, the explained variance was much higher for pond and floodplain biodiversity
441 compared with the values obtained for the river, lake and groundwater models.

442 Geo-climatic descriptors were significantly more influential than land use for the observed
443 biodiversity patterns. The variance partitioning scheme (BRT) revealed a consistently low
444 proportion of variance attributable to pure land use effects for all metrics (Figure 1).

445 Conversely, pure geo-climatic effects explained significantly more variance (Wilcoxon signed
446 rank test for paired samples: $P < 0.001$) across all ecosystems and organism groups and this
447 was more pronounced for river, lake and groundwater organisms and for pond amphibians
448 and invertebrates. Similarly, land use descriptors alone accounted for less than 3% of the
449 deviance (variation) in most GLM models (Table 2). Higher values ($>10\%$) were found only
450 for pond invertebrate and floodplain carabid beetle richness and for river invertebrate and

451 pond amphibian taxonomic distinctness. Both urban and agricultural land use performed
452 similarly in the models and no general pattern was obvious regardless of the biodiversity
453 metric considered.

454 The proportion of variance jointly attributable to both descriptor groups was equally high in
455 many cases and particularly pronounced with the floodplain and pond results (Figure 1). It
456 accounted for as much as 19–87% of the total variance in the floodplain biodiversity metrics
457 (ponds: 35–63%). It was also comparatively high for rivers (0.6–41%), but much lower for
458 lakes and groundwater (<12 and <10%, respectively for all metrics). Nevertheless, the joint
459 effects of land use and geo-climatic descriptors were significantly higher than the effects of
460 land use alone. The results suggest that both descriptor groups were intrinsically allied in
461 many models, which rendered the separation of unique effects difficult.

462 This, in part, was confirmed by significant interactions of geo-climatic and land use
463 descriptors, found for roughly a third of 33 GLM models and accounting for up to 17% of
464 model deviance (Table 3). The highest interactions (>10% explained deviance) were observed
465 for floodplain carabid beetles and molluscs and for pond amphibians, but the majority of
466 interaction terms accounted for less than 5% of the deviance in the models. Land use
467 interactions were strongest with longitude, latitude or annual precipitation, again highlighting
468 the intrinsic co-dependence between land use and geo-climatic descriptors.

469 However, the strong shared effects suggested by the variance partitioning scheme (Figure 1)
470 were not reflected by strong interactions in the GLMs (Table 3). Both were only moderately
471 positively correlated (Spearman's ρ : 0.455) (Figure 2).

472 We did not find a consistent decline in biodiversity in response to increasing land use
473 intensity (Table 2). More often than not the sign of the relationship was positive, i.e. the
474 biodiversity metrics increased with increasing percentages of arable and urban areas.

475 *3.2 Controlling the influence of geo-climatic descriptors by data sub-setting*

476 Latitude or longitude explained a considerable fraction of the variation in many biodiversity
477 metrics, regardless of the analytical approach applied. For example, in eleven out of 33 BRT
478 models, either latitude or longitude was the strongest geo-climatic descriptor, followed by
479 temperature (9 models), altitude (5), river catchment/lake surface area (3) and precipitation
480 (3) (Table 4). Temperature and precipitation, however, are also linked to latitude and
481 longitude at the European scale. By splitting the datasets along one of these (mostly)
482 geographical gradients the intention was to reduce the geographical extent of the derived data
483 subsets and hence to decrease the role of geo-climatic descriptors relative to the role of land
484 use in the data subsets.

485 Our findings are ambiguous and did not reveal a consistent pattern, neither with the
486 proportion of variance (deviance) explained by individual data subsets nor with the
487 geographical extent of the subsets. However, data sub-setting can control the analysis of land
488 use effects on freshwater biodiversity (Figure 3), as exemplified with lake phytoplankton and
489 pond invertebrate taxonomic distinctness (both split along the temperature gradient) and with
490 groundwater crustacean richness and rareness/endemicity (split along the evapotranspiration
491 and temperature gradient, respectively). In these cases, land use explained substantially more
492 deviance in the biodiversity metrics in both subsets.

493 In other cases, an increase was achieved in at least one subset, for example, with floodplain
494 mollusc richness. The subset split at annual precipitation ≤ 630 mm (Table 4) explained five
495 times the deviance in the full data and accounted for 50% of the metric's total deviance in this
496 subset. Likewise, the respective values doubled with floodplain carabid beetle and mollusc
497 rareness/endemicity and achieved explained deviances between 40 and nearly 50% for one
498 data subset (Figure 3, see Table 4 for the respective split points).

499 All but one of these data subsets were obtained by splits along gradients of actual or potential
500 evapo-transpiration, mean annual air temperature or altitude (Table 4). It appears the changes

501 observed in the deviance explained by land use (including interaction terms) when analysing
502 the data subsets were largely independent of the changes in the geographical (i.e. longitudinal
503 and latitudinal) extent within the subsets (Figure 4). Thus, although subsetting often also lead
504 to a decrease in the geographical extent of the subsets, the latitudinal or longitudinal splits of
505 the full data did not result in significant increases (or decreases) in the deviance explained by
506 the GLM models.

507 *3.3 Comparison of organism groups and biodiversity metrics*

508 In general, we were not able to detect strong consistent metric-driven differences across
509 ecosystems or organism groups. At ecosystem level, the high proportion of shared variance in
510 pond and floodplain diversity metrics was striking (mean \pm SD, ponds and floodplains: $54.5 \pm$
511 19% , rivers, lakes and groundwater: $11.6 \pm 10.4\%$).

512 On average, variance partitioning (BRT) explained most of the deviance in taxon richness
513 across all ecosystems and organism groups, followed by taxon rareness/endemicity and
514 taxonomic distinctness (significant only for richness *vs.* taxonomic distinctness). With
515 taxonomic distinctness, on average, 50% of the deviance remained unexplained. However,
516 with river invertebrates, pond amphibia and groundwater crustaceans taxonomic distinctness
517 performed comparatively well in the GLM models, especially in one of the data subsets
518 (Figure 3).

519 **4. Discussion**

520 This study presents biodiversity response patterns for different biodiversity metrics across
521 various taxa groups in aquatic and semi-aquatic ecosystem types. Given the broad European
522 scale of our study, we were required to use taxonomic data originating from national and
523 regional monitoring programmes. Where possible, these data were supplemented by data from
524 the scientific literature, adding another source of variability to the biological data. Although

525 there are limitations with the use of such data, particularly for biodiversity analysis, the
526 results we present show consistent response patterns. Therefore we are confident that the data
527 quality underlying our study was sufficient to draw the conclusions depicted below.

528

529 *4.1 Natural and anthropogenic land use effects on freshwater biodiversity*

530 Human land use, in particular urbanisation and intensified agriculture, are widely recognised
531 as major threats to freshwater biodiversity worldwide (MEA, 2005; Dudgeon *et al.*, 2006;
532 Vörösmarty *et al.*, 2010) and have been found to significantly impact the integrity of
533 freshwater systems (e.g. Allan, 2004; Feld *et al.*, 2011; Feld, 2013; Friberg 2014). Our
534 findings do not confirm this, but reveal a notably consistent pattern in terms of the weak
535 response of biodiversity to land use at the continental scale. The variance partitioning scheme
536 quantifies the role of land use in comparison to the natural descriptors of biodiversity and
537 reveals a low proportion of variation in biodiversity purely attributable to land use. Natural
538 geo-climatic descriptors are much better correlates of diversity, suggesting that both land uses
539 are less influential compared to the geo-climatic gradients at broad scales (Davies *et al.*,
540 2006). Among the natural descriptors considered, mean annual temperature, annual
541 precipitation, longitude, latitude and altitude form the most influential gradients in our data,
542 as confirmed by the split point analysis in BRT. This supports the assertion that energy and
543 climate are important in shaping diversity, as found by other studies (e.g. Davies *et al.*, 2007;
544 Mittelbach *et al.*, 2007; Field *et al.*, 2009), which will be highlighted further below.

545 The consistently strong shared effects of land use and geo-climatic descriptors (shared effects
546 were significantly higher than the pure land use effects) reveal agricultural and urban land
547 uses to be closely linked to geo-climatic conditions. Effects of both descriptor groups could
548 not be fully disentangled, which implies their consideration in tandem in macro-ecological
549 studies. In a similar study, Bruce *et al.* (2013) regressed fish diversity metrics in 1,632

550 European lakes against a selection of anthropogenic stressor variables and natural
551 (geographic) descriptors. They concluded that ‘geographical factors dominate over
552 anthropogenic pressures’. Although our results support this assertion, further specification is
553 required: geo-climatic factors not only dominate, but act in concert with land use. Hence,
554 broad-scale studies on environmental correlates of biodiversity must not overlook the shared
555 effects of natural and anthropogenic descriptors, which are consistently highlighted across
556 eleven organism groups and five ecosystem types in our study. This consistency in our
557 results, across aquatic and semi-aquatic ecosystems as well as across invertebrate and
558 vertebrate taxa groups is striking.

559 Although we were not able to further disentangle the shared effects, we found both shared
560 effects and interactions to be moderately positively correlated with each other. This suggests
561 interactions can explain strong shared effects, but not in all cases. Further investigation using
562 the spatial distribution of biodiversity (i.e. the potential spatial pattern) in Geographic
563 Weighted Regression may help locate regions where the shared effect of land use and geo-
564 climatic factors are particularly strong (Gouveia *et al.*, 2013) and thus help further disentangle
565 this linkage.

566 *4.2 Is data sub-setting the solution?*

567 The dominant role of geo-climatic descriptors (altitude, latitude and longitude) over human
568 impact at broad spatial scales suggests that human impact gradients are relatively short at such
569 broad scales (Davies *et al.*, 2006, Field *et al.*, 2009). We, therefore, split our data along the
570 major geo-climatic descriptor gradients and hypothesised that this data sub-setting would
571 enhance land use effects on biodiversity. Our results do not support the hypothesis, but reveal
572 rather inconsistent patterns, with both increasing as well as decreasing effects of land use
573 following data sub-setting. More often than not, at least one subset performed weaker than the
574 full data. This lets us conclude land use gradients remained comparatively short (i.e. weak)

575 also in the data subsets, although both land uses continued to span almost the entire possible
576 gradients (rivers: arable: 0 – >97% and artificial: 0 – >86% coverage) after sub-setting. Only
577 groundwater crustacean richness and rareness/endemicity showed increasing land use effects
578 in both subsets. We should note that these findings were also not linked to the different buffer
579 scales spanning several km for ponds up to 10,000 km² for groundwater and lake sites.

580 From this, we may conclude that the geographical extents of the regional subsets are still too
581 large to detect land use effects on biodiversity in our data. For example, the geographical split
582 of the river dataset at 51 °N and 6 °E results in a north-south expansion of 900 km in subset 1
583 (450 km, respectively for subset 2) and in a west-east expansion of 750 km (700 km,
584 respectively). However, our results imply that climatic gradients (temperature, precipitation)
585 influence freshwater biodiversity to a greater extent than geographical gradients (latitude,
586 longitude), thus indicating that climatic factors may dominate even at regional scales. This
587 again highlights a prevailing role of energy in shaping concurrent freshwater biodiversity
588 (Field et al., 2009; Tisseuil *et al.*, 2013; Bailly *et al.*, 2014) and explains the tendency for a
589 pronounced increase of effects of land use in at least one subset, when the split was along
590 climatic gradients. Hence, data sub-setting should aim to produce climatically more
591 homogeneous data in order to be able to analyse the human impact of land use on freshwater
592 biodiversity at the broad scale.

593 *4.3 The general response of freshwater biodiversity to land use*

594 A comparative analysis of measures of alpha diversity across ecosystems and organism
595 groups inevitably comes with potential methodological constraints. First, as much of our data
596 originate from national water quality monitoring schemes, field sampling methodologies
597 rarely fit a comprehensive biodiversity assessment. River samples, for example, often cover
598 only one season and only a limited area at a site (e.g. Feld *et al.*, 2013). Second, due to limited
599 determination capabilities, biodiversity may be difficult to estimate at the species level. Lake

600 phytoplankton, for example, covers a huge number of species, many of which usually present
601 in very low abundance (Carstensen *et al.*, 2005; Uusitalo *et al.*, 2013). Third, in particular
602 phytoplankton richness is strongly linked to sampling and counting methodology (Carstensen
603 *et al.*, 2005), while the determination is often restricted by the use of light microscopy of
604 preserved samples in routine monitoring schemes (Ojaveer *et al.*, 2010). These constraints
605 may influence our results and in part may limit the detection of stronger land use patterns.
606 Nevertheless, we believe that the concordance of patterns across ecosystems and organism
607 groups are striking and support the analytical approach followed in our study.

608 There is considerable evidence that urban (reviewed by Paul & Meyer, 2001) and agricultural
609 (reviewed by Allan, 2004, see also Feld *et al.*, 2013) land uses adversely affect the
610 biodiversity and integrity of lotic ecosystems. Likewise, pond macrophyte and invertebrate
611 richness are known to be negatively impacted by agriculture (Declerck *et al.*, 2006; Della
612 Bella & Laura, 2009) and pond amphibian and macrophyte richness by urbanisation (Akasaka
613 *et al.*, 2010; Hartel *et al.*, 2010). Similar adverse effects of human land use on freshwater
614 biodiversity are reported for lakes (Sala *et al.*, 2000; Hoffmann & Dodson, 2005; Bruçet *et*
615 *al.*, 2013) and obligate groundwater fauna (Malard *et al.*, 1996). We thus anticipated
616 pronounced negative effects of land use on freshwater biodiversity in this study.

617 Although we often found such negative effects, we also detected positive correlations
618 between the biodiversity metrics and coverage of both land use types. This was irrespective of
619 ecosystem type and organism group and has rarely been reported for aquatic ecosystems (but
620 see Hoffmann & Dodson, 2005). Due to the lack of nutrient data, we are not able to test the
621 response pattern along a more specific productivity gradient (Jeppesen *et al.*, 2000; Leibold,
622 1999). For both land use gradients, a unimodal response pattern, as evidenced by the study of
623 Hoffman & Dodson (2005) for lake zooplankton, was not evident in our data.

624 Species richness was, on average, the best performing metric in this study in terms of
625 response, followed by rareness/endemicity, while taxonomic distinctness was poorly
626 explained by the environmental descriptors in our analyses. Hence, our results partly support
627 the assumption that rare (and presumably sensitive) taxa respond to land use at the broad
628 scale, while taxonomic distinctness was a weak indicator of land use at this scale. It follows
629 that the taxonomic composition of communities changed along the land use gradients, while
630 the taxonomic relatedness of the community members remained relatively stable. This seems
631 contradictory to Feld *et al.* (2013) and related studies cited therein, but again could be
632 explained by the strong shared effects of climatic drivers of biodiversity and land use in our
633 broad-scale datasets. Seemingly, these not only drive freshwater species richness, but also
634 determine human land use patterns. Most likely, this applies to forms of intensive agriculture
635 (e.g. row-crops like maize), which are particularly dependant on suitable temperature and
636 precipitation regimes. Again this highlights the importance of broad-scale energy gradients in
637 macro-ecology (Hawkins *et al.*, 2003; Field *et al.*, 2009).

638 **5. Conclusions**

639 This study posits three major conclusions, with strong implications for future research on
640 freshwater biodiversity and its response to anthropogenic stressors at broad spatial scales:

- 641 1. At the European scale, natural geo-climatic descriptors, namely temperature, precipitation,
642 longitude and latitude, largely drive freshwater biodiversity. The same geo-climatic
643 descriptors are also strong determinants of human land use patterns, for example, of
644 agriculture and urbanisation. This results in considerable shared effects between natural and
645 human impact variables, which cannot be fully disentangled. Macro-ecological studies on
646 the effect of land use on biodiversity thus need to analyse both groups of descriptors
647 together. Simply using latitude and longitude as proxies of temperature and precipitation

648 thereby is not sufficient and more sophisticated spatial analytical methods are required (e.g.
649 Sharma *et al.* 2011).

650 2. Compared to land use (and probably also other human impacts), geo-climatic descriptors
651 form strong gradients in broad-scale datasets. Geo-climatically more homogeneous datasets
652 (i.e. subsets with less variation among natural explanatory variables) can help overcome the
653 dominance of natural gradients and may also provide stronger models explaining more
654 variance in the biological response variable. Sub-setting, however, does not translate to
655 simple geographical splits, for example into several regional subsets. Rather, sub-setting
656 should aim to cut (i.e. subdivide) the main geo-climatic gradient(s). Our study suggests a
657 split according to temperature and precipitation for several organism groups and freshwater
658 ecosystems.

659 3. Whole community-based biodiversity metrics, such as species richness and Shannon
660 diversity reveal contrasting responses to land use (and other anthropogenic stressor
661 gradients), likely to be caused by strong interactions with natural geo-climatic descriptors.
662 Measures of relative taxon rareness/endemicity and taxonomic distinctness (i.e.
663 phylogenetic diversity) did not perform better than measures of taxon richness and equal
664 distribution at the broad scale. Further studies should concentrate on the presence and
665 detection of species turnover along gradients of human impact. Also, measures that include
666 species identity might help detect human impact on freshwater biodiversity at the European
667 scale.

668 **6. Acknowledgements**

669 This study was supported by the European Commission through the BioFresh project: FP7-
670 ENV-2008, Contract no. 226874. We are thankful for to the following institutions and
671 organisations for providing data for this study: Mediterranean GIG (Data manager: Caridad de
672 Hoyos, CEDEX-CEH); Ministry of Agriculture, Natural Resources and Environment Cyprus,

673 Water Development Department (MANRE-WDD); water agencies and Institut national de
674 recherche en sciences et technologies pour environnement et agriculture (IRSTEA) and
675 ONEMA, France; Maria Moustaka, Aristotle University of Thessaloniki, Greece; Università
676 degli Studi di Sassari, Dipartimento di Scienze Botaniche, Ecologiche e Geologiche
677 (DiSBEG) Italy; Instituto da Água, I.P. (INAG), Portugal; Ministerio de Agricultura,
678 Alimentación y Medio Ambiente, Centro de Estudios Hidrográficos (CEDEX-CEH), Spain;
679 Central-Baltic GIG (Data manager: Ute Mischke, IGB); National Environmental Research
680 Institute, University of Aarhus, Denmark; Estonian Ministry of the Environment; Federal
681 Ministry for Agriculture, Forestry, Environment and Water Management (including all
682 Austrian Federal States); Austrian Federal Agency for Water Management; Institute of Water
683 Ecology, Fisheries and Lake Research Austria; Office of the Provincial Government of Lower
684 Austria; German Federal States of Bavaria, Hesse, Rhineland-Palatinate, North Rhine-
685 Westphalia, Lower Saxony, Schleswig-Holstein, Mecklenburg-Vorpommern, Saxony-Anhalt,
686 Saxony and Thuringia; Latvian Environment, Geology and Meteorology Centre; EPA
687 Lithuania; Rijkswaterstaat (RWS), the Netherlands; Institute of Environmental Protection—
688 National Research Institute, Poland; Scottish Environment Protection Agency (SEPA) and
689 Environment Agency for England & Wales (EA), UK; Eastern-Continental GIG (Data
690 manager: Gabor Borics, CER); Centre for Ecological Research, Hungarian Academy of
691 Sciences; Ministeriul Mediului și Pădurilor (MMP), Romania; Northern GIG (Data manager:
692 Geoff Phillips, EA); Finnish Environment Institute (SYKE); Environment Protection Agency
693 (EPA), Ireland; Norwegian Institute for Water Research (NIVA); Swedish University of
694 Agricultural Sciences (SLU). The following colleagues helped with the pond data: Lukács
695 Balázs András (Debrecen, HU), Jens Arle (Dessau, DE), Sandrine Angélibert (Allinges, FR),
696 Régis Céréghino (Toulouse, FR), Valentina Della Bella (Rome, IT), Carolin Meier
697 (Waiblingen, DE), Beat Oertli (Jussy, CH), Gwendolin Porst (Berlin, DE), Carl Sayer
698 (London, UK), Annika Schlusen (Essen, DE), Aneta Spyra (Katowice, PL), Sonja Stendera

699 (Solingen, DE), Nigel Willby (Stirling, UK) and Pablo Valverde (Bielefeld, DE). The
700 following colleagues contributed to the assembly of the European groundwater crustacean
701 data set: Maja Zagamajster (SI), Cene Fišer (SI), D. Galassi (IT), Fabio Stoch (IT) and Pierre
702 Marmonier (FR).

703 **7 References**

- 704 Akasaka M., Takamura N., Mitsuhashi H. & Kadono Y. (2010) Effects of land use on aquatic
705 macrophyte diversity and water quality of ponds. *Freshwater Biology*, **55**, 909–922.
- 706 Allan J.D. (2004). Landscapes and Riverscapes: The Influence of Land Use on Stream
707 Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 257–284.
- 708 Araújo M.B., Nogués-Bravo D., Diniz-Filho J.A.F., Haywood A.M., Valdes P.J. & Rahbek C.
709 (2008) Quaternary climate changes explain diversity among reptiles and amphibians.
710 *Ecography* **31**, 8–15.
- 711 Argillier, C., Caussé, S., Gevrey, M., Pédrón, S., De Bortoli, J., Brucet, S., Emmrich, M.,
712 Jeppesen, E., Lauridsen, T., Mehner, T., Olin, M., Rask, M., Volta, P., Winfield, I. J.,
713 Kelly, F., Krause, T., Palm, A. & Holmgren, K. (2013) Development of a fish-based index
714 to assess the eutrophication status of European lakes. *Hydrobiologia*, **704**, 193–211.
- 715 Baselga A., Gomez-Rodriguez C., Lobo J.M. (2012) Historical legacies in world amphibian
716 diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*
717 **7**, 323–341.
- 718 Bailly D., Cassemiro F.A.S., Agostinho C.S., Marques E.E. & Agostinho A.A. (2014) The
719 metabolic theory of ecology convincingly explains latitudinal diversity gradient of
720 Neotropical freshwater fish. *Ecology*, **95**, 553–562.
- 721 Birk S., Bonne W., Borja A., Brucet S., Courrat A., Pokane S., Solimini A., van de Bund W.,
722 Zampoukas N. & Hering D. (2012) Three hundred ways to assess Europe's surface waters:
723 An almost complete overview of biological methods to implement the Water Framework
724 Directive. *Ecological Indicators*, **18**, 31–41.
- 725 Böhmer J. (2012). Amphibienvorkommen und deren säurebedingten Laichschäden im
726 Nordschwarzwald und Odenwald in 2012 im Vergleich mit den Jahren 1987-1990, 1996
727 und 2002 (p. 51). BioForum GmbH, Kirchheim/Teck.
- 728 Brucet S., Pédrón S., Mehner T., Lauridsen T.L., Argillier C., Winfield I.J., Volta P.,
729 Emmrich M., Hesthagen T., Holmgren K., Benejam L., Kelly F., Krause T., Palm A., Rask

730 M. & Jeppesen E. (2013) Fish diversity in European lakes: geographical factors dominate
731 over anthropogenic pressures. *Freshwater Biology*, **58**, 1779–1793.

732 Carstensen J., Heiskanen A.S., Kauppila P., Neumann T., Schernewski G. & Gromizs S.
733 (2005) Developing Reference Conditions for Phytoplankton in the Baltic Coastal Waters.
734 *Part II: Examples of Reference Conditions Developed for the Baltic Sea*. Joint Research
735 Center, Technical Report, EUR 21582/EN/2, 35 pp.

736 Clapcott J.E., Collier K.J., Death R.G., Goodwin E.O., Harding J.S., Kelly D., Leathwick
737 J.R., Young R.G. (2012) Quantifying relationships between land-use gradients and
738 structural and functional indicators of stream ecological integrity. *Freshwater Biology*, **57**,
739 74–90.

740 Clarke K. & Warwick R. (1998) a taxonomic distinctness index and its statistical properties.
741 *Journal of Applied Ecology* **35**, 523–531.

742 Clarke K. & Warwick R. (1999) The taxonomic distinctness measure of biodiversity:
743 weighting of step lengths between hierarchical levels. *Marine Ecology Progress Series*
744 **184**, 21–29.

745 Crisp D.T., Laffan S., Linder H.P. & Monro A. (2001) Endemism in the Australian flora.
746 *Journal of Biogeography*, **28**, 183–198.

747 Cutler D. R., Edwards T.C., Beard K.H., Cutler A., Hess K.T., Gibson J. & Lawler J.J. (2007)
748 Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.

749 Dahm V., Hering D., Nemitz D., Graf W., Schmidt-Kloiber A., Leitner P., Melcher A., Feld
750 C.K. (2013) Effects of physico-chemistry, land use and hydromorphology on three riverine
751 organism groups: a comparative analysis with monitoring data from Germany and Austria.
752 *Hydrobiologia*, **704**, 389–415.

753 Davies R.G., Orme C.D.L., Olson V., Thomas G.H., Ross S.G., Ding T.-S., *et al.* (2006)
754 Human impacts and the global distribution of extinction risk. *Proceedings Biological*
755 *Sciences / The Royal Society*, **273**, 2127–2133.

756 Davies R.G., Orme C.D.L., Storch D., Olson V.A., Thomas G.H., Ross S.G., Ding T.S.,
757 Rasmussen P.C., Bennett P.M., Owens I.P.F., Blackburn T.M. & Gaston K.J. (2007)
758 Topography, energy and the global distribution of bird species richness. *Proceedings of the*
759 *Royal Society B: Biological Sciences*, **274**, 1189–1197.

760 De Meester L., Declerck S., Stoks R., Louette G., Van De Meutter F., De Bie T., Michels E.
761 & Brendonck L. (2005). Ponds and pools as model systems in conservation biology,
762 ecology and evolutionary biology. *Aquatic Conservation: Marine and Freshwater*
763 *Ecosystems*, **15**, 715–725.

- 764 Declerck S., De Bie T., Ercken D., Hampel H., Schrijvers S., Van Wichelen J., Gillard V.,
765 Mandiki R., Losson B., Bauwens D., Keijers S., Vyverman W., Goddeeris B., De Meester
766 L., Brendonck L. & Martens K. (2006). Ecological characteristics of small farmland ponds:
767 associations with land use practices at multiple spatial scales. *Biological Conservation*,
768 **131**, 523–532.
- 769 Della Bella V. & Mancini L. (2009). Freshwater diatom and macroinvertebrate diversity of
770 coastal permanent ponds along a gradient of human impact in a Mediterranean ecoregion.
771 *Hydrobiologia*, **634**, 25–41.
- 772 Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler D.J., Lévêque C.,
773 Naiman R.J., Prieur-Richard A.-H., Soto D., Stiassny M.L.J. & Sullivan C.A. (2006)
774 Freshwater biodiversity: importance, threats, status and conservation challenges.
775 *Biological Reviews*, **81**, 163.
- 776 Elith J., Leathwick J. R. & Hastie T. (2008). A working guide to boosted regression trees.
777 *Journal of Animal Ecology*, **77**, 802–813.
- 778 Ellenberg H., Weber H. E., Düll R., Wirth V., Werner W. & Paulißen D. (1992). Indicator
779 values of plants in Central Europe (Vol. 18, p. 258). Erich Goltze Verlag, Göttingen.
- 780 Evans K.L., Warren P.H. & Gaston K.J. (2005) Species–energy relationships at the
781 macroecological scale: a review of the mechanisms. *Biological Reviews* **80**, 1–25.
- 782 Feld, C.K. (2013) Response of three lotic assemblages to riparian and catchment-scale land
783 use: implications for designing catchment monitoring programmes. *Freshwater Biology*,
784 **58**, 715–729.
- 785 Feld C.K., Birk S., Bradley D.C., Hering D., Kail J., Marzin A., Melcher A., Nemitz D.,
786 Pedersen M.L., Pletterbauer F., Pont D., Verdonshot P.F.M. & Friberg N. (2011) Chapter
787 3 - From Natural to Degraded Rivers and Back Again: a Test of Restoration Ecology
788 Theory and Practice. *Advances in Ecological Research*, **44**, 119–209.
- 789 Feld C.K., de Bello F. & Dolédec S. (2013) Biodiversity of traits and species both show weak
790 responses to hydromorphological alteration in lowland river macroinvertebrates.
791 *Freshwater Biology*, **59**, 233–248.
- 792 Field R., Hawkins B. a., Cornell H. V., Currie D.J., Diniz-Filho J.A.F., Guégan J.-F., *et al.*
793 (2009) Spatial species-richness gradients across scales: a meta-analysis. *Journal of*
794 *Biogeography* **36**, 132–147.
- 795 Friberg, N. (2014) Impacts and indicators of change in lotic ecosystems. WIREs Water 2014.
796 doi: 10.1002/wat2.1040.

797 Gallardo B., Gascón S., Quintana X., Comín F.A. (2011) How to choose a biodiversity
798 indicator - Redundancy and complementarity of biodiversity metrics in a freshwater
799 ecosystem. *Ecological Indicators*, **11**, 1177–1184.

800 Gouveia S.F., Hortal J., Cassemiro F. a. S., Rangel T.F. & Diniz-Filho J.A.F. (2013)
801 Nonstationary effects of productivity, seasonality, and historical climate changes on global
802 amphibian diversity. *Ecography* **36**, 104–113.

803 Guégan J.-F., Lek S. & Oberdorff T. (1998) Energy availability and habitat heterogeneity
804 predict global riverine fish diversity. *Nature*, **391**, 382–384.

805 Harding J.S., Benfield E.F., Bolstad P.V., Helfman G.S. & Jones E.B.D. III. (1998) Stream
806 biodiversity: The ghost of land use past. *Proceedings of the National Academy of Science*
807 *USA*, **95**, 14843–14847.

808 Hartel T., Schweiger O., Öllerer K., Cogălniceanu D. & Arntzen J. W. (2010). Amphibian
809 distribution in a traditionally managed rural landscape of Eastern Europe: Probing the
810 effect of landscape composition. *Biological Conservation*, **143**, 1118–1124.

811 Hawkins B.A., Field R., Cornell H. V., Currie D.J., Guégan J.-F., Kaufman D.M., Kerr J.T.,
812 Mittelbach G.G., Oberdorff T., O'Brien E.M., Porter E.E. & Turner J.R.G (2003) Energy,
813 water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

814 Heino, J. (2013) The importance of metacommunity ecology for environmental assessment
815 research in the freshwater realm. *Biological Reviews*, **88**, 166–178.

816 Heino, J. (2011) A macroecological perspective of diversity patterns in the freshwater realm.
817 *Freshwater Biology*, **56**, 1703–1722.

818 Heino, J., Soininen, J., Lappalainen, J. & Virtanen, R. (2005) The relationship between
819 species richness and taxonomic distinctness in freshwater organisms. *Limnology and*
820 *Oceanography*, **50**, 978–986.

821 Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. (2005). Very high resolution
822 interpolated climate surfaces for global land areas. *International Journal of Climatology*,
823 **25**, 1965–1978.

824 Hijmans R.J., Phillips S., Leathwick J. & Elith J. (2013) dismo: Species Distribution
825 Modeling. R Package Version 0.8-11. Available at: [http://CRAN.R-](http://CRAN.R-project.org/package=dismo)
826 [project.org/package=dismo](http://CRAN.R-project.org/package=dismo).

827 Hof C., Brändle M. & Brandl R. (2008) Latitudinal variation of diversity in European
828 freshwater animals is not concordant across habitat types. *Global Ecology and*
829 *Biogeography*, **17**, 539–546.

- 830 Hoffmann M.D. & Dodson S.I. (2005) Land use, primary productivity, and lake area as
831 descriptors of zooplankton diversity. *Ecology*, **86**, 255–261.
- 832 Homes V., Hering D. & Reich M. (1999). The distribution and macrofauna of ponds in
833 stretches of an alpine floodplain differently impacted by hydrological engineering.
834 *Regulated Rivers: Research & Management*, **417**, 405–417.
- 835 Hortal J., Diniz-Filho J.A.F., Bini L.M., Rodríguez M.Á., Baselga A., Nogués-Bravo D., *et al.*
836 (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung
837 beetles. *Ecology Letters*, **14**, 741–748.
- 838 Illies J. (1978) *Limnofauna Europaea*. 2. Auflage. Gustav Fischer Verlag, New York,
839 Stuttgart. 532 pp.
- 840 Irz, P., De Bortoli, J., Michonneau, F., Whittier, T. R., Oberdorff, T. & Argillier, C. (2008)
841 Controlling for natural variability in assessing the response of fish metrics to human
842 pressures for lakes in north-east USA. *Aquatic Conservation: Marine and Freshwater*
843 *Ecosystems*, **18**, 633-646.
- 844 Jeppesen E., Jensen J.P., Sondergaard, M., Lauridsen T. & Landkildehus F. (2000) Trophic
845 structure, species richness and biodiversity in Danish lakes: Changes along a phosphorus
846 gradient. *Freshwater Biology*, **45**, 201–218.
- 847 Jeppesen E., Meerhoff M., Holmgren K., Gonzalez-Bergonzoni I., Teixeira-de Mello F.,
848 Declerck A.A.J. *et al.* (2010) Impacts of climate warming on lake fish community structure
849 and potential effects on ecosystem function. *Hydrobiologia*, **646**, 73–90.
- 850 Jeppesen E., Mehner T., Winfield I.J., Kangur K., Sarvala J., Gerdeaux D., *et al.* (2012)
851 Impacts of climate warming on the long-term dynamics of key fish species in 24 Euro-
852 pean lakes. *Hydrobiologia* **694**, 1–3.
- 853 Legendre P. & Legendre L. (1998) *Numerical Ecology*, 2nd edition. Elsevier, Amsterdam.
- 854 Leibold M.A. (1999) Biodiversity and nutrient enrichment in pond plankton communities.
855 *Evolutionary Ecology Research*, **1**, 73–95.
- 856 Leprieur F., Tedesco P., Hugueny B., Beauchard O., Dürr H.H., Brosse S., *et al.* (2011)
857 Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures
858 of past climate changes. *Ecology Letters*, **14**, 325–334.
- 859 Linder H.P. (2001) Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of*
860 *Biogeography*, **28**, 169–182.
- 861 Loh J. & Wackernagel M. (ed.) (2004) *Living Planet Report 2004*. World-Wide Fund for
862 Nature International (WWF), Global Footprint Network, UNEP World Conservation
863 Monitoring Centre, Gland, Switzerland.

864 MacArthur R.H. & Wilson E.O (1967) *The theory of island biogeography*. Princeton
865 University Press, Princeton, New Jersey, USA.

866 Malard F., Mathieu j., Reygrobellet J.-L. & Lafont M. (1996) Biomonitoring groundwater
867 contamination: application to a karst area in Southern France. *Aquatic Sciences*, **58**, 158–
868 187.

869 MEA (Millennium Ecosystem Assessment) (2005). *Ecosystems and Human Well-being:*
870 *Synthesis*. Island Press, Washington, DC.

871 Mittelbach G.G., Schemske D.W., Cornell H.V, Allen A.P., Brown J.M., Bush M.B., *et al.*
872 (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and
873 biogeography. *Ecology Letters*, **10**, 315–331.

874 Moe S.J., Schmidt-Kloiber A., Dudley B.J. & Hering D. (2013). The WISER way of
875 organising ecological data from European rivers, lakes, transitional and coastal waters.
876 *Hydrobiologia*, **704**, 11–28.

877 Nagorskaya L., Moroz M., Laeno T., Veznovetz V., Moller Pillot H., Dijkstra K.D.B. &
878 Reemer M. (2002). Macrofauna in floodplain pools and dead branches of the Pripyat river,
879 Minsk. Institute of Zoology, National Academy of Science Belarus, p. 158.

880 Oberdorff T., Tedesco P., Hugueny B., Leprieur F., Beauchard O., Brosse S., *et al.* (2011)
881 Global and Regional Patterns in Riverine Fish Species Richness: A Review. *International*
882 *Journal of Ecology*, **2011**, 1–12.

883 Oertli B., Auderset Joye D., Castella E., Juge R., Lehmann A. & Lachavanne J.-B. (2005)
884 PLOCH: a standardized method for sampling and assessing the biodiversity in ponds.
885 *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 665–679.

886 Ojaveer H., Jaanus A., MacKenzie B.R., Martin G., Olenin S., Radziejewska T., Telesh I.,
887 Zettler M.L. & Zaiko A. (2010) Status of Biodiversity in the Baltic Sea. *PLoS ONE*, **5**,
888 e12467.

889 Paul M.J. & Meyer J.L. (2001) Streams in the urban landscape. *Annual Review of Ecology*
890 *and Systematics*, **32**, 333–365.

891 Pearson, R.G. & Boyero, L. (2009) Gradients in regional diversity of freshwater taxa. *Journal*
892 *of the North American Benthological Society*, **28**, 504–514.

893 Phillips G., Lyche-Solheim, A., Skjelbred B., Mischke U., Drakare S., Free G., Järvinen M.,
894 de Hoyos C., Morabito G., Poikane S. & Carvalho L. (2013) A phytoplankton trophic
895 index to assess the status of lakes for the Water Framework Directive. *Hydrobiologia*, **704**,
896 75–95.

897 Poff N.L. (1997). Landscape filters and species traits: towards mechanistic understanding and
898 prediction in stream ecology. *Journal of the North American Benthological Society*, **16**,
899 391–409.

900 R Development Core Team (2013). R: A Language and Environment for Statistical
901 Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at:
902 <http://www.R-project.org>. (Version: 2.15.3)

903 Reyjol Y., Hugueny B., Pont D., Bianco P.G., Beier U., Caiola N., Casals F. Cowx I.G.,
904 Economou A., Ferreira M. T., Haidvogel G., Noble R., de Sostoa A., Vigneron T. &
905 Virbickas T. (2007). Patterns in species richness and endemism of European freshwater
906 fish. *Global Ecology and Biogeography*, **16**, 65–75.

907 Ridgeway G. (2013) gbm: Generalized Boosted Regression Models R Package Version 2.0-8.
908 Available at: <http://CRAN.Rproject.org/package=gbm>.

909 Sala O.E., Chapin F.S. III, Armesto J.J., Berlow E., Bloomfield J., Dirzo R., Huber-Sanwald
910 E., Huenneke L.F., Jackson R.B., Kinzig A., Leemans R., Lodge D.M., Mooney H.A.,
911 Oesterheld M., Poff N.L., Sykes M.T., Walker B.H., Walker M. & Wall D.H. (2000)
912 Global Biodiversity Scenarios for the Year 2100. *Science*, **287**, 1770–1774.

913 Sayer C.D., Andrews K., Shilland E., Edmonds N., Edmonds-Brown R., Patmore I., Emson
914 D. & Axmacher J.A. (2012) The role of pond management for biodiversity conservation in
915 an agricultural landscape. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **22**,
916 626–638.

917 Scheuerell M.D. & Schindler D.E. (2004) Changes in the spatial distribution of fishes in lakes
918 along a residential development gradient. *Ecosystems*, **7**, 98–106.

919 Schindler D.W. (2006) Recent advances in the understanding and management of
920 eutrophication. *Limnology and Oceanography*, **51**, 356–363.

921 Sobkowiak S. (2003) Vergleichende Untersuchungen zur Makroinvertebratenfauna von
922 Auengewässern – unter besonderer Berücksichtigung der Kinzig-Aue (Hessen). Diploma
923 thesis, University of Duisburg-Essen, Germany.

924 Stendera S., Adrian R., Bonada N., Cañedo-Argüelles M., Hugueny B., Januschke K.,
925 Pletterbauer F., & Hering D. (2012) Drivers and stressors of freshwater biodiversity
926 patterns across different ecosystems and scales: a review. *Hydrobiologia*, **696**, 1–28.

927 Tisseuil C., Cornu J.-F., Beauchard O., Brosse S., Darwall W., Holland R., Hugueny B.,
928 Tedesco P.A. & Oberdorff T. (2013) Global diversity patterns and cross-taxa convergence
929 in freshwater systems. *Journal of Animal Ecology*, **82**, 365–376.

930 Uusitalo L., Fleming-Lehtinen V., Hällfors H., Jaanus A., Hällfors S. & London L. (2013) A
931 novel approach for estimating phytoplankton biodiversity. *ICES Journal of Marine*
932 *Science*. doi:10.1093/icesjms/fss198

933 Venables W.N. & Ripley B.D. (2002) *Modern Applied Statistics with S (MASS)*. 4th Ed.
934 Springer, New York. Available: <http://cran.r-project.org/web/packages/MASS/index.html>.
935 (Accessed on 1 April 2014).

936 Vinson, M.R. & Hawkins, C.P. (2003) Broad-scale geographical patterns in local stream
937 insect genera richness. *Ecography*, **26**, 751–767.

938 Vinson, M.R. & Hawkins, C.P. (1998) Biodiversity of stream insects: variation at local, basin,
939 and regional scales. *Annual Review of Entomology*, **43**, 271–293.

940 Vörösmarty C.J., McIntyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P., Glidden
941 S., Bunn S.E., Sullivan C.A., Reidy Liermann C. & Davies P.M. (2010) Global threats to
942 human water security and river biodiversity. *Nature*, **467**, 555–561.

943 Warton D.I. & Hui F.K.C. (2011) The arcsine is asinine: the analysis of proportions in
944 ecology. *Ecology* **92**, 3–10.

945 Wetzel R. (2001) *Limnology—Lake and River Ecosystems*. Academic Press, New York.

946 Wright D.H. (1983). Species-energy theory: an extension of species-area theory. *Oikos*, **41**,
947 496–506.

948 Zgmajster M., Eme D., Fiser C., Galassi D., Marmonier P., Stoch F., Cornu J.F. & Malard F.
949 (2014) Geographic variation in range size and beta diversity of groundwater crustaceans:
950 insights from habitats with low thermal seasonality. *Global Ecology and Biogeography*,
951 DOI: 10.1111/geb.12200.

952 Zomer R.J, Trabucco A., Bossio D.A., van Straaten, O. & Verchot L.V. (2008) Climate
953 Change Mitigation: A Spatial Analysis of Global Land Suitability for Clean Development
954 Mechanism Afforestation and Reforestation. *Agriculture, Ecosystems and Environment*,
955 **126**, 67–80.

956 Zuur A.F., Ieno E.N. & Smith G.M. (2007) *Analysing ecological data*. Springer, New York,
957 672 pp.

958

Tables

Table 1: Main characteristics of the land use and biological data.

	Land use data source	Area [km ²]	Shape of area	Organism groups (No. of sites)	No. of taxa (determination level)
Rivers	CORINE 2006	variable (entire catchment upstream of a site)	irregular	Fish (590) Invertebrates (1,221) Macrophytes (651)	66 (species) 564 (genus) 234 (species)
Floodplains	CORINE 2006	78.5 km ²	circle around site (radius = 5 km)	Carabidae (132) Mollusca (81) Vegetation (352)	301 (species) 185 (species) 1,205 (species)
Ponds	CORINE 2006	4.9 km ²	circle around pond (radius = 125 m)	Amphibia (148) Invertebrates (189) Macrophytes (392)	34 (species) 416 (species, genus) 320 (species)
Lakes	CORINE 2006	10,000 km ²	100 x 100 km grid	Phytoplankton (836)	970 (species, genus)
Groundwater	GlobCover	10,000 km ²	100 x 100 km grid	Crustacea (21,700 entries for 494 grid cells)	1,570 (species)

Table 2: Matrix of strength and direction of biodiversity metrics in response to urban and agricultural land use across all ecosystem types and organism groups. Response strengths and direction ('+' positive, '-' negative relationship) are according to the highest deviance explained by land use (without interaction terms) in the GLM models using the complete datasets: $>|10\%| = +++/- --$; $>|5\%| = ++/- -$; $>|3\%| = +/-$; $\leq|3\%| = 0$.

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Rivers	Fishes	0	+	0	++	0	0
Rivers	Invertebrates	0	-	0	0	++	+++
Rivers	Macrophytes	--	0	0	--	0	0
Lakes	Phytoplankton	0	0	0	0	0	0
Ponds	Amphibia	0	+	0	0	+++	--
Ponds	Coleoptera/Odonata/Gastropoda	0	+++	++	0	0	0
Ponds	Macrophytes	0	++	0	++	-	0
Floodplains	Carabidae	---	--	++	-	0	0
Floodplains	Mollusca	0	+	0	0	0	--
Floodplains	Macrophytes	--	0	0	0	0	0
Groundwater	Crustacea	+	0	0	0	0	--

Table 3: Percent deviance explained by significant interaction terms including land use in the GLM models based on the complete datasets. If more than one interaction was significant, the total deviance explained by all interactions is provided. Geo-climatic descriptor(s) interacting with land use are listed in brackets; area = catchment size; lat = latitude; lon = longitude; ppt = annual precipitation; temp= mean annual air temperature; pet = potential evapo-transpiration; hab = habitat diversity.

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Rivers	Fishes		1.4 (area)				
Rivers	Invertebrates						
Rivers	Macrophytes				1.5 (lat)		
Lakes	Phytoplankton						
Ponds	Amphibia	14.4 (lon)			3.1 (lon)	11.3 (lon)	
Ponds	Coleoptera/Odonata/Gastropoda			5.1 (lat, ppt, temp)	6.6 (lat)	2.7 (ppt)	
Ponds	Macrophytes	1.8 (ppt)	2.3 (lat, ppt)	3.7 (pet)			1.2 (lat)
Floodplains	Carabidae		11.1 (ppt)	9.5 (ppt)			

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Floodplains	Mollusca	4.1 (lon)		17 (ppt)			
Floodplains	Macrophytes				3.7 (temp)		1.9 (lon)
Groundwater	Crustacea	1 (hab)					

Table 4: Split points used to generate two data subsets for each combination of ecosystem type, organism group and biodiversity metric. Split points were identified using the partial dependence plots provided by the Boosted Regression Tree models, but were modified in order to achieve a more balanced sample size in both subsets. For clarity, subset 1 always encompasses the samples \leq split point and subset 2 the samples $>$ the split point. See text for details.

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Rivers	Fish richness	Catchment size	500 km ²	516	74
Rivers	Fish rareness	Catchment size	500 km ²	516	74
Rivers	Fish taxonomic distinctness	Catchment size	500 km ²	516	74
Rivers	Invertebrate richness	Latitude	51° N	639	582
Rivers	Invertebrate rareness	Latitude	51° N	639	582
Rivers	Invertebrate taxonomic distinctness	Latitude	51° N	639	582
Rivers	Macrophyte richness	Longitude	6° E	96	555
Rivers	Macrophyte rareness	Latitude	51° N	292	359
Rivers	Macrophyte taxonomic distinctness	Latitude	51° N	191	303
Lakes	Phytoplankton richness	Mean annual air temperature	6 °C	192	644
Lakes	Phytoplankton rareness	Mean annual air temperature	7.7 °C	315	521
Lakes	Phytoplankton taxonomic distinctness	Mean annual air temperature	9.3 °C	655	181
Ponds	Amphibia richness	Mean annual air temperature	8.8 °C	110	38

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Ponds	Amphibia rareness	Ecoregion	4 (yes/no)	alpine: 84	non-alpine: 64
Ponds	Amphibia taxonomic distinctness	Mean annual air temperature	8.9 °C	89	35
Ponds	Coleoptera/Odonata/Gastropoda richness	Annual precipitation	992 mm	119	58
Ponds	Coleoptera/Odonata/Gastropoda rareness	Latitude	48 °N	109	62
Ponds	Coleoptera/Odonata/Gastropoda taxonomic distinctness	Mean annual air temperature	8.8 °C	53	124
Ponds	Macrophyte richness	Latitude	49 °N	338	263
Ponds	Macrophyte rareness	Latitude	49 °N	338	263
Ponds	Macrophyte taxonomic distinctness	Annual precipitation	839 mm	238	327
Floodplains	Carabidae richness	Annual mean air temperature	9.9 °C	62	70
Floodplains	Carabidae rareness	Altitude	37 m a.s.l.	62	70
Floodplains	Carabidae taxonomic distinctness	Altitude	55 m a.s.l.	76	55
Floodplains	Mollusca richness	Annual precipitation	630 mm	51	30
Floodplains	Mollusca rareness	Longitude	16.5 °E	32	47
Floodplains	Mollusca taxonomic distinctness	Longitude	12.3 °E	32	47

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Floodplains	Macrophyte richness	Annual mean air temperature	9.9 °C	170	182
Floodplains	Macrophyte rareness	Altitude	49 m a.s.l.	150	202
Floodplains	Macrophyte taxonomic distinctness	Altitude	19 m a.s.l.	124	198
Groundwater	Crustacea richness	Evapotranspiration (AET)	600 mm	406	120
Groundwater	Crustacea endemism	Mean annual air temperature	10.9 °C	134	256
Groundwater	Crustacea taxonomic distinctness	Altitude	462 m a.s.l.	217	121

Figure captions

Figure 1: Variance partitioning scheme using four biodiversity metrics and eleven organism groups sampled in five ecosystem types. Each plot displays the pure and shared proportions of variance explained and unexplained by land use and geo-climatic variables in the Boosted Regression Tree analyses. NA = Shannon diversity cannot be computed with presence/absence data.

Figure 2: Percent deviance explained by significant interaction terms in GLM against percent variance explained by shared effects (BRT). A GAM smoother was overlaid the scatter plot to highlight the relationship of both variables.

Figure 3: Proportion of deviance explained by land use and interactions with land use in the GLM models using three biodiversity metrics calculated for eleven organism groups. Each model run was repeated using the full dataset (filled symbol) and two data subsets (empty symbols). Data subsets were generated separately for each biodiversity metric and based on the split points identified by Boosted Regression Tree analysis for the strongest geo-climatic environmental descriptor variable in each model.

Figure 4: Changes in the proportion of deviance explained by land use (GLM models, absolute values) against percent range of latitude and longitude covered by data subsets 1 and 2 in comparison to the range of the full dataset. High percent values on the x-axis indicate a higher resemblance of latitude and longitude gradients to those of the full dataset. For the definition of subsets 1 and 2, see Table 4.

Figures

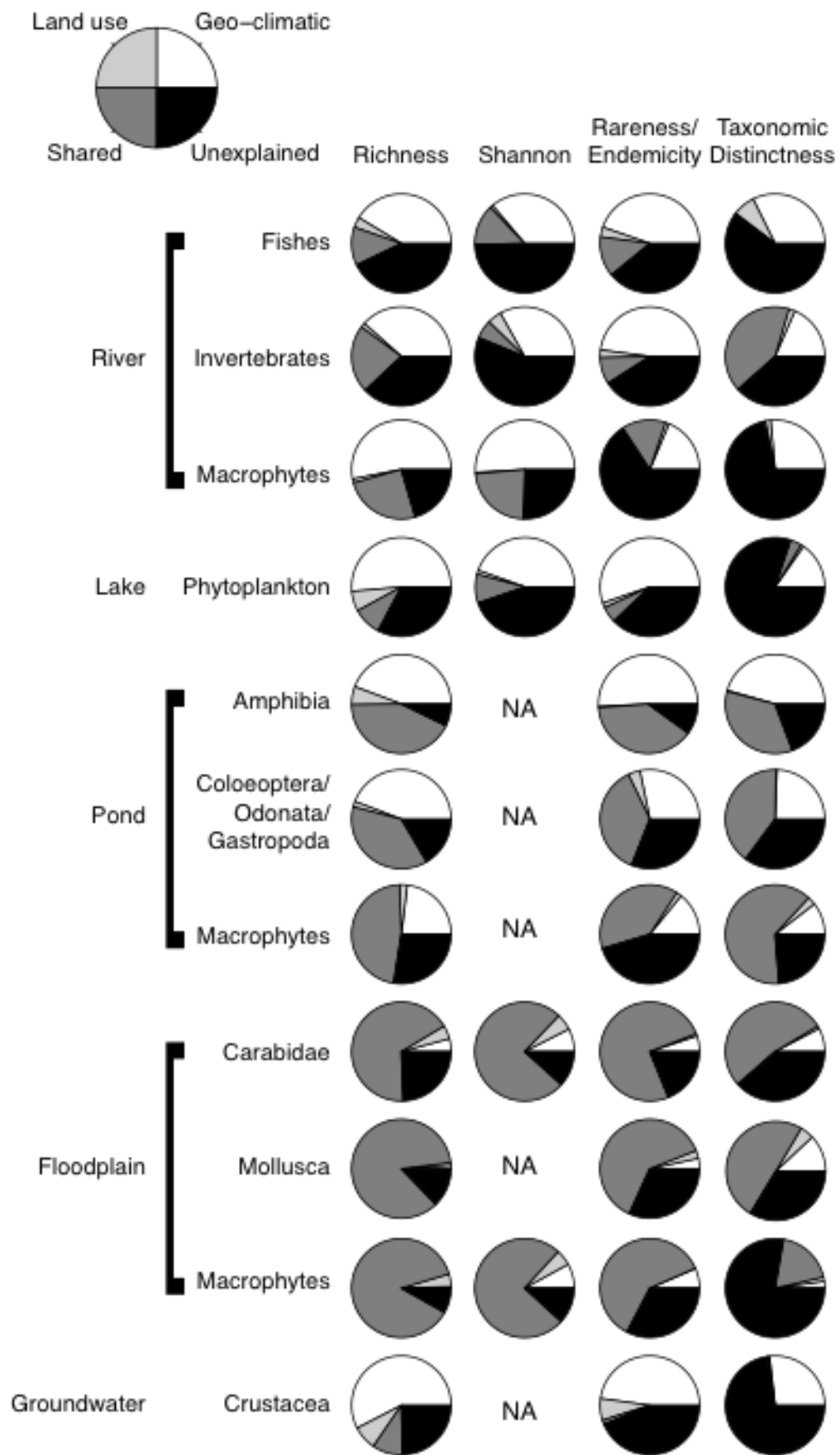


Figure 1

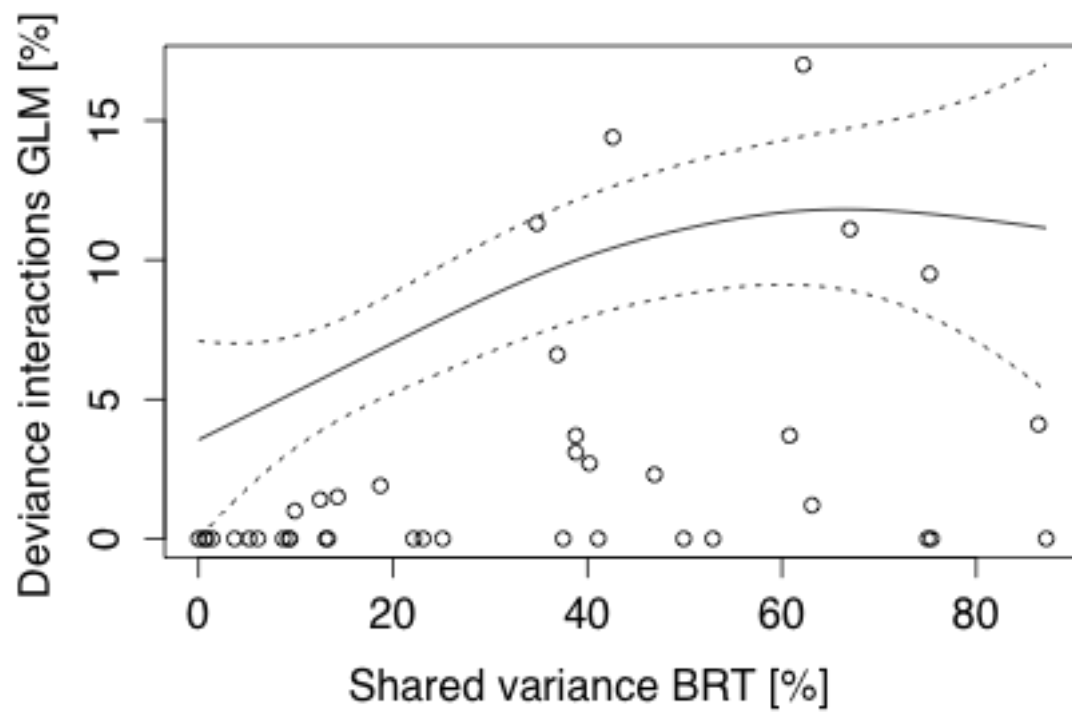


Figure 2

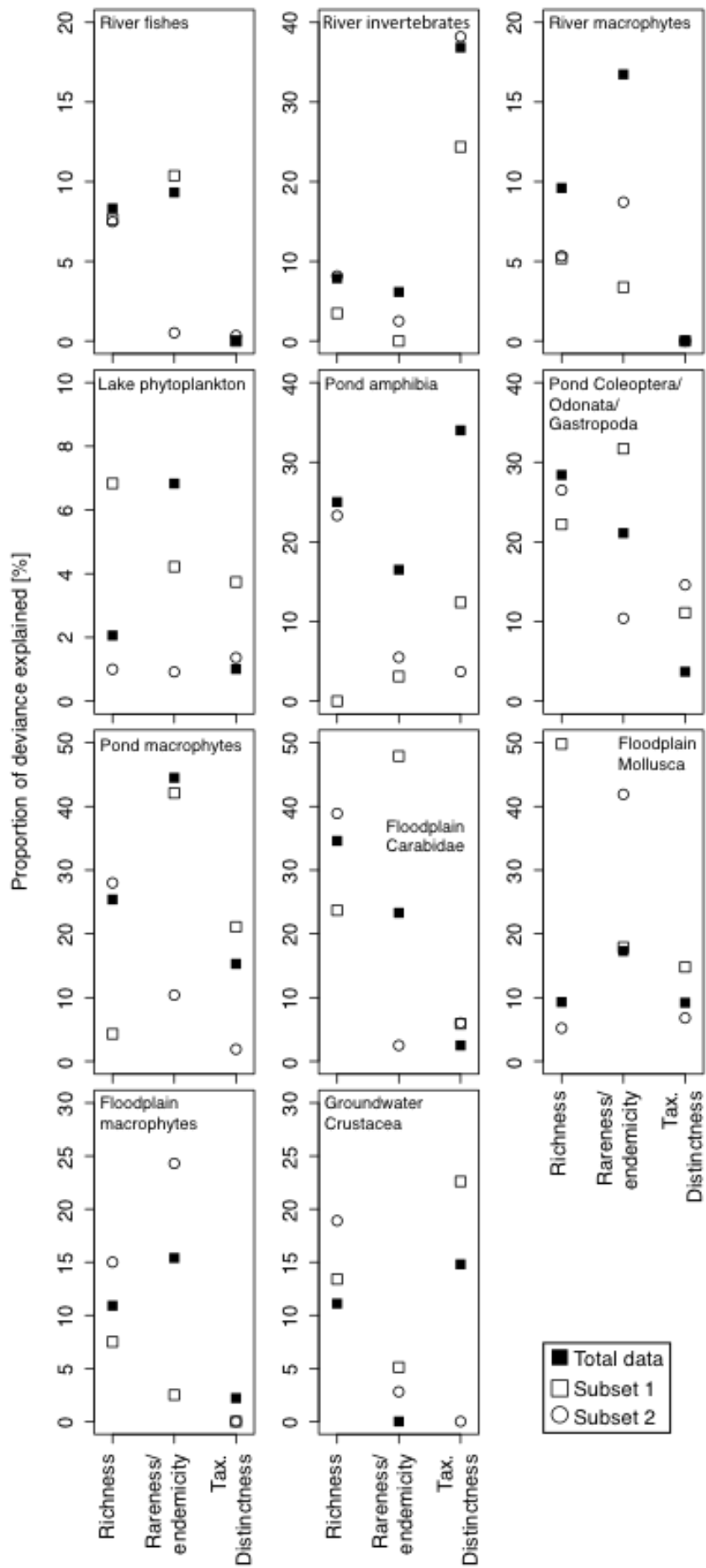


Figure 3

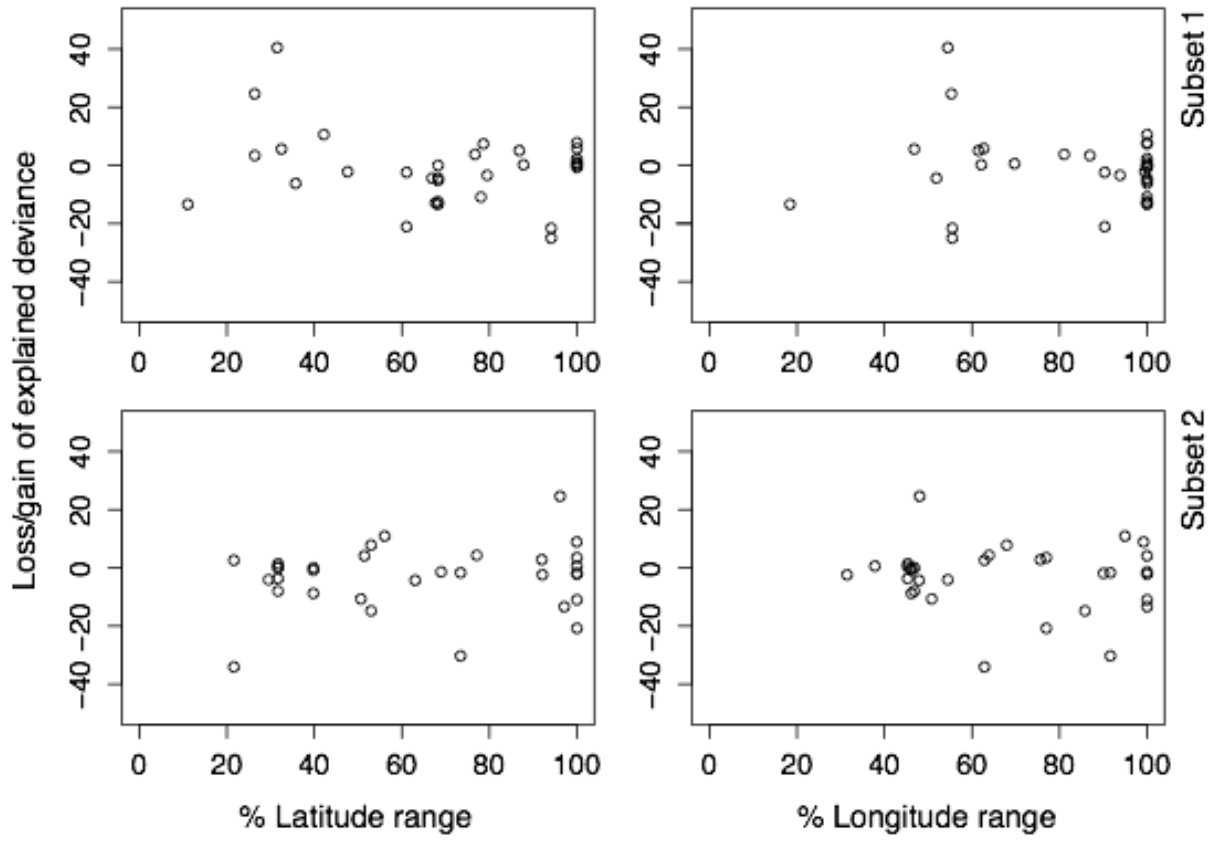


Figure 4

Supplementary Material

Table S1: Non-collinear geo-climatic variables used as descriptors in the Boosted Regression Trees and Generalised Linear Modelling.

Descriptors	Lakes	Rivers	Ponds	Wetlands/ floodplains	Groundwater
Longitude (°E)	x	x	x	x	x
Latitude (°N)		x			
Altitude (m a.s.l.)	x		x	x	x
Mean annual temperature (°C)	x	x	x	x	x
Annual precipitation (mm)	x	x	x	x	x
Catchment size (km ²)		x			
Actual evapo- transpiration (mm)					x
Potential evapo- transpiration (mm)			x		
Surface area (km ²)	x				