1   Stream diatom assemblages as environmental indicators

2   – a cross-regional assessment

3

4   Virpi Pajunen[1], Maria Kahlert[2] and Janne Soininen[1]

5

6   *[1]Department of Geosciences and Geography, P.O. BOX 64, FIN-00014 University of*

7   *Helsinki, Finland*

8   *[2]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural*

9   *Sciences, BOX 7050, 75007 Uppsala, Sweden*

10

11

12

13

14

15

16

17

18

19   Correspondence: Virpi Pajunen, Department of Geosciences and Geography, P.O. Box 64,

20   FI-00014 University of Helsinki, Finland. E-mail: virpi.pajunen@helsinki.fi.

21

22

23   **Running title:** Diatoms as environmental indicators

24   **Key words:** bioindicators, biomonitoring, predictive modelling, stream diatoms

Abstract

Ongoing climatic change and anthropogenic pressure highlight the importance of reliable

assessment of the ecological status of freshwaters. Bioindicators are used widely in

ecological assessments as biotic assemblages reflect the environmental conditions in current

ecosystems. However, the robustness of bioindicators relies on the transferability of indices

and models outside the regions they were derived from. We tested the reliability of stream

diatom assemblages as indicators of water chemistry and climatic factors in a cross-regional

assessment by developing a predictive model with diatom assemblage data from Sweden and

using it to model stream conditions in Finland. The inference models and predictions were

performed using the Boosted Regression Trees (BRT) method, separately in species and

genus levels. The predictive performance of the calibration models in Sweden were good or

moderate for both water chemistry and climatic variables, both at species and genus levels.

The most important climatic (growing degree days, $r^2 = 0.57$) and water chemistry variables

(pH, $r^2 = 0.65$; and total phosphorus (TP), $r^2 = 0.52$) could be inferred from diatom

assemblages relatively well. However, predictive performances of the cross-regional models

were low ($r^2 \leq 0.13$). Nevertheless, water chemistry variables, conductivity ($r^2 = 0.13$) and TP

($r^2 = 0.12$), were predicted the best. The most important diatom indicators for climatic and

environmental variables varied between Sweden and Finland. The study showed that diatom

assemblages can be robust indicators of water chemistry and climatic variables within the

region where the inference models are calibrated. However, their indicator ability may be

weak between regions. The reason for the low transferability of the diatom inference models

may stem from between-region differences in species realized niches, species pools and/or

ecosystems, local adaptation or species identification. Hence, models should only be used

with caution in geographical contexts other than the one where they were developed. The

calibration data should cover as large geographical area as possible to give reliable

50   predictions when applied in smaller regions.

51

1. Introduction

53   Water is an essential element for life. However, ongoing climate change and increasing

54   anthropogenic pressures affect the quality of all kinds of aquatic ecosystems (Foley et al.

55   2005, Wrona et al. 2006). Freshwater ecosystems are especially threatened (Heino et al.

56   2009). In order to monitor and predict the changes caused by the changing climate and land

57   use, robust methods are necessary. Although assessments of water quality using bioindicators

58   are conducted around the world, many regions do not have a robust system for water quality

59   assessments (Taylor et al. 2007, Chen et al. 2016, Tan et al. 2017). Furthermore, the potential

60   effect of climatic factors on bioindicators is typically not integrated in existing biotic indices,

61   which are calibrated mainly with water chemistry. This may hinder the reliability of these

62   indices under rapidly changing climatic conditions and in cross-regional assessments.

63   Bioindicators are widely used to infer the ecological status of ecosystems. The robustness of

64   bioindicators relies on the known environmental preferences of indicator species (Smol &

65   Stoermer 2010). Depending on the life-cycles of the studied taxa, biological assemblages

66   reflect the conditions for a variable period of time rather than reflecting conditions only

67   during a snapshot like water chemistry sampling does (Sandin & Verdonschot 2006).

68   Additionally, communities reflect the overall conditions, i. e. the mixture of chemical

69   compounds, physical environment and biological interactions providing a cost-efficient

70   method for environmental surveys. In freshwaters and especially in fluvial ecosystems, the

71   most commonly used bioindicators are macroinvertebrates and microscopic algae, especially

72   diatoms (Resh 2008). Diatoms (Bacillariophyceae) are microscopic unicellular algae, living

73   in almost any kind of waters, and can be identified typically to species and variety level by

74 their unique siliceous cell wall (Round et al. 1990). Environmental preferences of many

75 diatom species are known relatively well (Smol & Stoermer 2010) and many species have a

76 specific range of tolerance for pH or nutrient enrichment, for instance (e.g. van Dam et al.

77 1994). Currently used diatom indices, such as IPS (Cemagref 1982) or ACID (Andrén &

78 Jarlman 2008), are based on these known tolerances.

79 Traditionally, indices are built using simple methods such as weighted averaging (Cemagref

80 1982, Kelly & Whitton 1995, Coste et al. 2009). A disadvantage of weighted averaging is that

81 it assumes a unimodal response of species to environment, which often is not the case in

82 nature. A more novel method would be a machine learning approach such as Boosted

83 Regression Trees (BRT), which takes into account more complex responses, interactions

84 between predictors and thresholds above or below which species may occur (Elith et al.

85 2008).

86 There are other problems with the accuracy of diatom indices (e.g. Besse-Lototskaya et al.

87 2011). As a growing number of countries are developing their own indices for water quality

88 assessment, many still use indices developed in some other geographical region (Taylor et al.

89 2007, Chen et al. 2016, Tan et al. 2017). Recently, it has been shown that the observed optima

90 and tolerances of diatom species might vary between geographical regions (Rimet et al. 2007,

91 Soininen et al. 2019; but see Bennett et al. 2010). This observed variation among species'

92 responses indicates that the indices developed in one region may not be transferable across

93 regions. This is disappointing since the usage of diatoms as indicators (e.g. of water

94 chemistry) relies on the concept that diatom distributions are ubiquitous and restricted only

95 by local environmental variables to similar degree across regions (Smol & Stoermer 2010).

96 Although, this might be the case for many cosmopolitan species, a number of studies suggest

97 that endemism, local adaptation and dispersal limitation occur among diatom species

98   (Vanormelingen et al. 2008, Jüttner et al. 2010). In fact, ultimate factors, such as climate,

99   geology and land use, can in some circumstances explain more variation in diatom

100 communities and species distributions than local environmental factors (Pajunen et al.

101 2016a).

102 We studied the usefulness of stream diatom assemblages as indicators of water chemistry and

103 climatic factors across two boreal regions. We developed a predictive model using Swedish

104 diatom data and used it to model stream conditions in a neighboring country, Finland. A

105 recent study has shown that diatom assemblages can be good indicators for both local

106 environmental and climatic factors (Pajunen et al. 2016b); however, the applicability of such

107 predictive model have not yet been tested outside the regions they were calibrated.

108

109 **2. Materials and methods**

110 *2.1. Diatom and physico-chemical data*

111 Two extensive data sets of stream diatom assemblages were obtained from the neighboring

112 countries Sweden (55° – 70° N, 10° – 25° E) and Finland (60° – 70° N, 20° – 32° E) in order

113 to conduct a cross-regional assessment of the reliability of bioindicators (Fig. 1). The

114 Swedish data set consisted of 571 stream sites (Table 1). The Swedish samples were collected

115 for the national and several regional routine sampling programs across Sweden during

116 autumn. Diatoms were sampled by brushing stones with a toothbrush, according to the

117 recommendations of EU standards ((European Committee for Standardization 2014a; 2014b).

118 At least five replicate, pebble-to-cobble sized stones (5-15 cm) were selected randomly from

119 five to 10 cross-stream transects. They were brushed and the diatom suspension was placed in

120 a small plastic bottle and preserved in ethanol (70%). In the laboratory, organic materials

121   were cleaned from diatom samples by wet combustion with hydrogen peroxide and mounted

122   in Naphrax or Dirax. At least 400 valves per sample were counted using phase contrast light

123   microscopy (magnification 1000x) and identified to the lowest taxonomic level possible

124   using the literature in the Swedish standard (Jarlman et al. 2016). Identifications were done

125   by experts who had previously harmonized their identification routines within the Nordic-

126   Baltic Network for Benthic Algae in Freshwater (NorBAF; Kahlert & Albert 2005), meaning

127   that their counts were as similar as counted replicates, and diatom names were harmonized

128   according to conventions adopted by NorBAF (Kahlert et al. 2009). Sampling for water

129   chemistry and other local parameters are part of those routine sampling programs, and water

130   chemistry values consist of either spot values taken simultaneously with diatom samples or

131   monthly mean values. Both diatom and other data are stored in the national database

132   (Department of Aquatic Sciences and Assessment SLU 2010).

133   The Finnish data set comprised 227 stream sites collected between 1986 and 2004. The sites

134   were located in the five ecoregions in Finland and covered broad gradients in water chemistry

135   (Soininen et al. 2004) (Table 1). All sampling was conducted during low flow conditions in

136   July and August. At the majority of the sites, water samples were taken simultaneously with

137   the diatom samples and analyzed for total phosphorus (TP), pH, conductivity and water color.

138   For less than 20% of the sites, water chemistry data were taken from the national water

139   quality database (Finnish Environment Institute 2019), using results from the nearest

140   sampling occasion in space and time. Current velocity and stream width were measured and

141   shading by riparian vegetation was assessed at each site along 10 transects perpendicular to

142   the flow covering the whole sampling area. Diatom samples were collected and prepared

143   following standard protocols as in Sweden, with the exception that combustion was done with

144   acid (HNO_3:H_2SO_4; 2:1) (European Committee for Standardization 2014a; 2014b). For each

145   sample, 250–500 valves were counted using phase contrast light microscopy. Species were

146 identified to the lowest level possible according to Krammer & Lange-Bertalot (1986-1991)

147 and Lange-Bertalot & Metzeltin (1996) by two analysts who had harmonized their species

148 identification routines by working together.

149 The Swedish and Finnish data sets were carefully harmonized by diatom experts following

150 the NorBAF conventions, involving the merging of subspecies and matching of both data

151 sets. After harmonization the datasets comprised 878 species in Sweden and 213 in Finland

152 (see Appendix A, Table A.1). All taxa not found in both data sets (n=699) were removed

153 keeping the relative abundances of each species as in original full data. Thus, the summed

154 relative abundances in the Swedish data set ranged from 27.5 to 100 % (md 82.2 %), and in

155 the Finnish data set from 61.3 to 100 % (md 91.3 %). Additionally, diatom genera data were

156 generated by merging this new reduced species data into diatom genera. In the end, both

157 Swedish and Finnish data sets contained the same 196 species and 65 genera.

158 *2.2. Climatic data*

159 Climate data (averages for the years 1979–2013) were obtained from the CHELSA climate

160 data base (Karger et al. 2017a, 2017b). The data set includes high resolution (30 arc sec ≈ 1

161 km) climatic data for the earth land surface areas, including monthly mean temperature and

162 precipitation patterns. Frost degree days (FDD, defined as temperature < 0 ˚C), growing

163 degree days (GDD, defined as temperature > 5 ˚C), mean annual precipitation (MAP) and

164 water balance (WAB) were defined for each Finnish and Swedish stream site from the

165 climatic data rasters. FDD indicates the length and severity of the frozen winter period,

166 whereas GDD can be seen as a proxy for temperature and overall productivity of the stream

167 and its catchment. MAP and WAB indicate the amount of atmospheric water supply to the

168 whole catchment area, potential run-off and flow regime.

169 *2.3. Data analyses*

170 Covariance between the local and climatic variables was assessed using Spearman's rank

171 correlation separately for both countries. Climatic and environmental data exhibited strong

172 collinearities ($\geq |0.70|$) between some variables in both data sets (Table A.2). Coordinates (i.

173 e. latitude and longitude), FDD and MAP were later excluded from the predictions due to the

174 strong correlations with other variables, such as GDD and WAB. GDD and WAB were

175 selected as they reflect the overall climatic conditions in the growing season. Overlying

176 histograms (Fig. A.1) show that the ranges of climatic and environmental variables were

177 much larger in the Swedish data set than in the Finnish data set. To avoid extrapolation,

178 Swedish data were chosen to be used as the calibration data to predict Finnish environmental

179 conditions. Principal component analysis (PCA) was used to further explore the

180 environmental data of both data sets (Fig. A.2, Table A.3). Environmental variation was

181 much larger in the Swedish data set than in Finnish data set. This supports the decision to use

182 Swedish data as the calibration data for the predictions.

183

184 The most important variables for diatom assemblages in the data sets were investigated by

185 using RDA, both at species and genus level. The strongly correlated variables (coordinates,

186 FDD and MAP) were removed prior the analyses. To further visualize the relationships

187 between diatom assemblages and the climatic and environmental variables, we performed

188 non-metric multidimensional scaling (NMDS) for all data sets (Figs. A.3 and A.4). PCA,

189 RDA and NMDS analyses were performed in R (version 3.2.2; R Development Core Team,

190 2016) applying VEGAN package (Oksanen et al. 2015).

191

192 *2.4. Calibration and inference models*

193 The prediction models were conducted by using Swedish data as calibration data and Boosted

194 Regression Trees (BRT) as a modelling approach. BRT is a machine learning method where

8

195   predictions with minimized loss function (such as deviance) are created by a boosting

196   technique in which a sequence of simple regression trees is progressively grown fitting one

197   tree at a time to the sequence (Friedman et al. 2000, De'ath, 2007, Elith et al., 2008). All

198   models were fitted using R software (version 3.2.2; R Development Core Team 2016).

199   Models were conducted separately for the species and genera data. In model calibration,

200   GDD, WAB, pH, TP, conductivity and water color were set as response variables and the

201   diatom abundance of 196 taxa or 65 genera from 571 sites as predictors. The BRT model was

202   fitted using functions from the gbm package version 2.1.3 (Ridgeway 2017) based on

203   Friedman's (2001) gradient boosting machine. The interaction depth in the model was set to

204   6. The learning rate, determining the contribution of each tree to the growing model, was set

205   to 0.005 and the maximum number of trees was set to 1000. A Gaussian distribution of errors

206   was used to model the six variables. Leave-one-out cross-validation (LOO) was used to

207   assess the performance of the models. Model performance was estimated by the coefficient of

208   determination (r^2) and the root-mean-square error of prediction (RMSEP). The relative

209   importance of the predictor variables was first estimated according to Friedman (2001) and

210   then scaled to sum up to 100. The higher value a predictor variable gets, the stronger its

211   influence on the response variable. Additional calibration models were performed using

212   Finnish data in order to estimate the relative importance of species and genera in Finland.

213   Finally, the Swedish calibration models were used to predict climatic and environmental data

214   for Finnish data.

215

216   Alternative models were conducted to guard against potential differences among calibration

217   models, datasets and used modelling techniques. Firstly, the performance of calibration

218   models using Finnish dataset and their predictive ability in Swedish data were tested.

219   Secondly, a subset of Swedish dataset (n = 396 sites, subsampled to match the ranges in

220   Finnish data, see Appendix C, Table C.1) was used as calibration data to infer climatic and

221   environmental variables for Finnish data. Thirdly, a weighted averaging method was used to

222   repeat the original calibration and inference models (Swedish data as calibration dataset (n =

223   571 sites) for sites in Finland). The WA models were fitted using functions from the RIOJA

224   package version 0.9-21 (Juggins 2019) in R. The performance of the calibration models was

225   assessed using leave-one-out cross-validation and model performance was estimated by the

226   coefficient of determination (r^2) and the root-mean-square error of prediction (RMSEP).

227

228   **3. Results**

229   PCA analyses showed much larger variation in environmental and climatic variables in

230   Swedish than in Finnish data set (Fig. A.2, Table A.3). The first two components collectively

231   explained 67% of the variation in Swedish data and 66% in Finnish data. In Swedish data,

232   latitude, longitude and FDD had high positive and GDD a negative loading on the first

233   component, while MAP and WAB had high positive and pH a negative loading on the second

234   component. In Finnish data, latitude and FDD had high negative and GDD a positive loading

235   on the first component, whereas pH had a high negative loading on the second component.

236   Based on the RDAs, pH, GDD and TP were the most important factors explaining variation

237   in diatom species composition of the set of variables tested in Sweden, whereas GDD,

238   conductivity and TP were the most important factors in Finland (Fig. 2, see also Appendix B,

239   Table B.1). The most important factors for diatom genera did not differ from those for

240   species, except that in Finnish data, where TP was slightly more important than conductivity

241   (Fig. B.1, Table B.2).

242

243 Within Sweden, the predictive performance of the calibration models was moderate ($r^2 \geq$

244 0.27) for all variables (Table 2, Figs. 3 and B.2). The highest coefficients of determination

245 between observed and diatom inferred values were for pH ($r^2 = 0.65$ with species data and r^2

246 = 0.71 with genus data), GDD ($r^2 = 0.57$ and $r^2 = 0.53$) and TP ($r^2 = 0.52$ with both species

247 and genus).

248

249 When predicting from Sweden to Finland, the predictive performance of all the models was

250 low ($r^2 = 0.002$ to 0.13) (Table 2, Figs. 4 and B.3), although evaluation of the prediction

251 models showed moderate performances for the most important variables, GDD, pH, TP and

252 conductivity ($r^2 = 0.31$ and 0.33). Nevertheless, water chemistry variables conductivity ($r^2 =$

253 0.13 (species) and 0.11 (genera)) and TP ($r^2 = 0.12$ and 0.10) were best predicted. On

254 average, environmental variables could be predicted slightly better using species than genus

255 models. None of the alternative models showed significant improvement in the model

256 performances (Tables C.1, C.3 and C.4). The predictive performances of the alternative

257 models showed moderate to good performance for Swedish data and poor to moderate

258 performance for Finnish data. The level of model performance did not considerably change

259 whether a country was used as calibration dataset or being inferred for, i.e. calibration models

260 using Finnish dataset had poor performance in Finland yet they could infer for Swedish data

261 reasonably well.

262

263 The most important species and genera as predictors for each variable varied between

264 Sweden and Finland (Figs. 5 and B.4). In the top ten of the most important species, only two

265 (for color, conductivity and TP) or three (for GDD, WAB and pH) species were the same in

266 both countries (Fig. 5). The ten most important genera included four (for conductivity, TP

267   and GDD), five (for pH and WAB) or six (for color) genera that were among the ten most

268   important in both countries (Fig. B.4).

269

270   **4. Discussion**

271   The performances of the calibration models within Sweden demonstrated that the most

272   important climatic and water chemistry variables can be inferred from diatom assemblages

273   relatively well, both with species and genus level taxonomy. The predictive ability for pH and

274   TP showed similar or better accuracy than diatom inference models conducted previously in

275   streams in other regions (e.g. USA (pH and TP; Pan et al. 1996), Finland (pH; Soininen &

276   Niemelä 2002), Finland (TP; Pajunen et al. 2016b)). GDD was also well-predicted within the

277   calibration set when compared with inference models created earlier in Finland with the same

278   method (Pajunen et al. 2016b). These results indicate that diatom assemblages are robust

279   indicators of both local environmental and climatic variables, when the inference models are

280   being used in the same geographical region where they have been calibrated.

281   However, we found only low transferability of our predictive models from Sweden to Finland

282   for both local environmental and climatic variables. Although some differences among

283   species responses were expected between the two boreal neighboring countries, such a poor

284   predictive performance for Finland was surprising even if the performances of the calibration

285   models were good or moderate. This result is especially unexpected concerning pH and TP,

286   which are variables typically well-predicted by diatom assemblages (Pan et al. 1996,

287   Soininen & Niemelä 2002, Pajunen et al. 2016b). The failure of the cross-regional predictions

288   could be caused by several reasons: 1) between-region differences in species realized niche

289   brought about by competition or stressors, 2) different species pools and ecosystems in the

290   study regions, or 3) identification differences and the effect of data selection. Below, we

291   discuss these aspects in more detail.

292   *4.1. Fundamental vs realized niche*

293   The poor predictive performance could result from different realized niches for species in

294   Sweden and Finland. A species has a fundamental ecological niche, which is often described

295   as an optima and range of tolerance for environmental conditions (Hutchinson 1957).

296   However, in the presence of competition or environmental stressors, a species is forced to use

297   only a part of its range of tolerance (Hutchinson 1959, Soberón & Peterson 2005, Peterson

298   2011). As the size and composition of regional species pool, environment and the dominance

299   of competitive species may differ among geographical regions, the realized niches of species

300   may vary among different regions (Soberón & Peterson 2005, Pearman et al 2008). This

301   creates a problem in species based inference models as species' fundamental niches can

302   seldom, or never, be fully covered in natural environments (Jackson & Overpeck 2000,

303   Jiménez et al. 2019). In contrast, the observed niches of diatom species in nature are typically

304   the realized niches in a certain geographical area. Thus, the inference models are most likely

305   to perform better in the region they have been calibrated than outside the calibration region.

306   The differences in the realized niches between the regions influence the optima and

307   tolerances of species used in diatom indices. In fact, many studies have found a discrepancy

308   in the indicator values of some diatoms species. For example, Potapova & Charles (2007)

309   pointed out that some studies have reported low nutrient optimum for *Ulnaria ulna* (Kelly &

310   Whitton 1995, Soininen & Niemelä 2002), yet some have found it to indicate high-nutrient

311   conditions (Rott et al. 1997). Such differences in species' responses to environmental

312   variables between regions may be caused by the variable degree of biotic interactions, local

313   adaptation, or different environmental conditions, but then, these differences may also stem

314   from identification problems (see section 4.3). Also, the ranges of environmental variables in

315   the datasets influence the derived values of species' optima and tolerances. Models calibrated

316   from larger geographical areas may perform better as they include wider ranges of

317   environmental variables and may thus be able to detect the species' realized niche more

318   reliably. Additionally, mean values of temporal environmental measurements can reflect more

319   robustly the environmental conditions affecting species than snapshot water chemistry

320   measurements. On the other hand, environmental interactions, interspecific interactions or

321   even microevolution may increase the noise in such datasets, which again could reduce the

322   performance of models based on these datasets.

323   *4.2 Different species pools and ecosystems*

324   Differences in the species pools and environmental conditions between the regions where the

325   calibration and inferred data sets originated may decrease the predictive ability of diatom

326   inference models. Here, Swedish and Finnish data sets consisted of the same 196 species,

327   which generally represented the majority of individuals in sites in both countries (median

328   coverage in Finland was 91.3% and in Sweden 82.2% of the individuals). However, in some

329   sites the species assemblages were dominated by species not occurring in the original data set

330   of the neighboring country. This indicates a difference in species pools between these two

331   neighboring countries, yet also, the variation in species assemblages may stem from

332   environmental differences (Table 1, Fig. A.4) between Sweden and Finland. Based on the

333   RDA results, the most influential water chemistry variable for Finnish diatom communities

334   was conductivity, whereas in Sweden, pH was the most important environmental factor. This

335   shows that the main drivers of diatom community composition differ between the two

336   countries indicating different environmental settings and limiting resources. The poor

337   performance of Finnish calibration models (Table C.1) compared to earlier studies using the

338   same data set with full species data (Pajunen et al. 2016b) indicate that the subset of Finnish

14

339   diatom assemblages used in this study lack the ability to indicate environmental conditions in

340   Finland.

341   The differences between species pools in Sweden and Finland could also reflect some

342   dispersal limitation due to dispersal barriers (e.g. the Baltic Sea). Additionally, both countries

343   may contain some peculiar ecosystems, such as heavily polluted, acidic or brownificated

344   streams, not found in the other country (or at least not covered in the data set), and these

345   ecosystems may cover such environmental conditions and gradients that harbor specialist

346   taxa. For instance, the relative abundances of the indicator species used in the models covered

347   only 27.5% out of the total abundance (500 cells) in the site with the highest water color (535

348   mg Pt L^{-1}) in Sweden. This indicates that the dominant species in this site is especially

349   tolerant of high water color, and its absence in Finnish data set may be due to the smaller

350   range of water color in the Finnish data set.

351   Generally, one reason for the models' poor ability to infer environmental variables may be the

352   lack of specialized species in the data sets. Therefore, the indication ability of the

353   assemblages is weak due to the lack of joint indicator species between the data sets. Many

354   good indicator species are rare as they have a clear optima for a certain environmental

355   variable and a narrow range of tolerance. However, integrating these rare species into

356   inference models is problematic as their low occurrence can cause noise and instability to the

357   models. In this study, rare species were not systematically left out, since the species data from

358   Sweden and Finland were matched to consist of all the species present in both data sets. Thus,

359   keeping all the species in the models would probably enhance the predictive ability of the

360   calibration set, but not the inference model.

361   *4.3 Problems in identification or data selection*

362   In addition to cryptic species, diatom models can also suffer from other problems in species

15

363   identification (Kahlert et al. 2012, Werner et al. 2016). The identification process is very

364   detailed and the accurate interpretation of complex structures is in the end subjective.

365   Therefore, the probability of differences in identification is greater when more analysts have

366   participated in the species identification. However, such differences can be minimized by

367   identifying diatoms to genus level only. In this study, the differences between the predictive

368   ability of species and genera based models were only minor, indicating that our

369   harmonization efforts had been successful and species identification did not seem to cause

370   major problems in this study. The similar performance of the two models may also indicate

371   that the whole diatom genera responded to environment quite similarly and/or that the use of

372   genus level in identification overall decreased the between-analyst differences in species

373   identification, and models performed well even if the sensitivity of genus level to indicate

374   environmental conditions may be somewhat lower than the species level.

375   Finally, we note that the inconsistencies in observed species optima and the range of

376   tolerance between different geographical areas can be caused by differences in ranges of

377   environmental variables in each data set. It has been shown that the relative importance of an

378   environmental variable for diatom community depends on its range of variation in the data set

379   (Potapova & Charles 2002). Moreover, in streams, environmental conditions can fluctuate

380   rapidly, and thus, a snapshot measurement of a water chemistry variable may not reflect the

381   actual conditions faced by the biota accurately enough. This may have partly caused the poor

382   predictive ability of environmental variables in Finland as the water chemistry values in

383   Finnish dataset were based on snapshot measurements whereas the Swedish environmental

384   data consisted of mean values from different sampling occasions. Furthermore, the

385   unidirectional flow may increase the potential mass effects, i.e., a constant dispersal of

386   species from high-quality patches to sink habitats (Pulliam 1988, Terui et al. 2014, Jamoneau

387   et al. 2018). Therefore, species may occur at a site that does not match their preferences

388   because of such metapopulation dynamics.

389

390   **5. Conclusions**

391   The present study shows that diatom assemblages are reliable indicators of water chemistry

392   and climatic variables within the region where the inference models are calibrated, yet their

393   indicator ability shows low transferability across regions. The underlying reasons for the

394   weak performance of the diatom inference models may be related to between-region

395   differences in species realized niche, species pools and/or ecosystems, local adaptation or

396   species identification. These findings highlight the importance of being careful when using

397   models based on regional datasets in a different geographical context.

398

405

406   **Appendices**

407   Supplementary data associated with this article can be found, in the online version:

408   *Appendix A*: The results of data analyses, including species data, covariance matrixes,

409   overlying histograms, NMDS and PCA.

410   *Appendix B*: The results of RDAs, and the calibration and inference models for genus data.

411   *Appendix C*: The results of alternative calibration and inference models.

412

413   **References**

414   Allan J.D., Castillo M.M., 2007. Stream Ecology: Structure and Function on Running Waters,

415   second ed. Springer, Dordrecht.

416   Andren, C., Jarlman, A., 2008. Benthic diatoms as indicators of acidity in streams.

417   Fundamental and Applied Limnology, 173: 237–253. https://doi.org/10.1127/1863-

418   9135/2008/0173-0237.

419   Bennett, J. R., Cumming, B. F., Ginn, B. K., J. P. Smol., 2010. Broad-scale environmental

420   response and niche conservatism in lacustrine diatom communities. Global Ecology and

421   Biogeography 19: 724–732. https://doi.org/10.1111/j.1466-8238.2010.00548.x.

422   Besse-Lototskaya, A., Verdonschot, P. F. M., Coste, M., Van de Vijver, B., 2011. Evaluation

423   of European diatom trophic indices. Ecological Indicators 11: 456–467.

424   https://doi.org/10.1016/j.ecolind.2010.06.017.

425   Cemagref, 1982. Etude des méthodes biologiques d´appréciation quantitative de la qualité des

426   eaux. Q.E. Lyon-A.F.Bassion Rhône-Méditeranée-Corse, pp. 218.

427   Chen, X., Zhou, W., Pickett, S. T. A., Li, W., Han, L., Ren, Y., 2016. Diatoms are better

428   indicators of urban conditions: a case study in Beijing, China. Ecological Indicators 60:

429   265–274. https://doi.org/10.1016/j.ecolind.2015.06.039.

430   Coste, M., Boutry, S., Tison-Rosebery, J., Delmas, F., 2009. Improvements of the Biological

431   Diatom Index (BDI): description and efficiency of the new version (BDI-2006).

432   Ecological Indicators 9: 621–650. https://doi.org/10.1016/j.ecolind.2008.06.003.

433   van Dam, H., Mertens, A., Sinkeldam, J., 1994. A coded checklist and ecological indicator

434   values of freshwater diatoms from the Netherlands. Netherlands Journal of Aquatic

435   Ecology 28: 117–133. https://doi.org/10.1007/BF02334251.

436 De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243–251.

437   https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.

438 Department of Aquatic Sciences and Assessment SLU 2010. Data host for inland waters.

439   Available at: http://www.slu.se/en/departments/aquatic-sciences-assessment/ (last

440   accessed 9 June 2015).

441 Elith, J., Leathwick, J. R., Hastie, T., 2008. A working guide to boosted regression trees.

442   Journal of Animal Ecology 77: 802–813. https://doi.org/10.1111/j.1365-

443   2656.2008.01390.x.

444 European Committee for Standardization, 2014a. EN 13946:2014 Water quality – Guidance

445   for the routine sampling and preparation of benthic diatoms from rivers and lakes. C.F.R.

446 European Committee for Standardization, 2014b. EN 14407:2014 Water quality – Guidance

447   for the identification and enumeration of benthic diatom samples from rivers and lakes.

448   C.F.R.

449 Finnish Environment Institute, 2019. Open information. Available at:

450   http://www.syke.fi/avointieto (accessed September 2019).

451 Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S.,

452   Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A.,

453   Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, C., Snyder, P. K.,

454   2005. Global consequences of land use. Science 309: 570–574.

455   https://doi.org/10.1126/science.1111772.

456 Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. Annals

457   of Statistics 29:1189–1232. https://doi.org/10.1214/aos/1016218223.

458 Friedman, J., Hastie, T., Tibshirani, R, 2000. Additive logistic regression: a statistical view of

459   boosting. Annals of Statistics 28: 337–407.

460   Heino, J., Virkkala, R., Toivonen, H., 2009. Climate change and freshwater biodiversity:

461   detected patterns, future trends and adaptations in northern regions. Biological Reviews

462   84: 39–54. https://doi.org/10.1111/j.1469-185X.2008.00060.x.

463   Hutchinson, G. E., 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative

464   Biology 22: 415–427.

465   Hutchinson, G. E., 1959. Homage to Santa Rosalia or why are there so many kinds of

466   animals? The American Naturalist 93: 145–159.

467   Jackson, S. T., Overpeck, J. T., 2000. Responses of plant populations and communities to

468   environmental changes of the late Quaternary. Paleobiology 26: 194–220.

469   Jamoneau, A., Passy, S. I., Soininen, J., Leboucher, T., Tison-Rosebery, J. 2018. Beta

470   diversity of diatom species and ecological guilds: Response to environmental and spatial

471   mechanisms along the stream watercourse. Freshwater Biology 63: 62–73.

472   Jarlman, A., Kahlert, M., Sundberg, I., Herlitz, E., 2016. Påväxt i sjöar och vattendrag −

473   kiselalgsanalys. Version 3:2: 2016-01-20. Available at:

474   https://www.havochvatten.se/download/18.2daa1277152c4afdb3058f99/1455617201913

475   /undersokstyp-pavaxt-i-sjoar-och-vattendrag-kiselalgsananlys.pdf (accessed March

476   2019).

477   Jiménez, L., Soberón, J., Christen, J. A., Soto, D., 2019. On the problem of modeling a

478   fundamental niche from occurrence data. Ecological Modelling 397: 74–83.

479   https://doi.org/10.1016/j.ecolmodel.2019.01.020.

480   Juggins, S. 2019. Package ´rioja`. Available at: https://cran.r-

481   project.org/web/packages/rioja/rioja.pdf (accessed October 2019).

482   Jüttner, I., Chimonides, P. D. J., Cox, E. J., 2010. Ecology and biogeography of Himalayan

483   diatoms: distribution along gradients of altitude, stream habitat and water chemistry.

484   Fundamental and Applied Limnology 177: 293–311. https://doi.org/10.1127/1863-

485   9135/2010/0177-0293.

486   Kahlert, M., Albert, R.-L., 2005. NorBAF - The Nordic-Baltic Network for Benthic Algae in

487   Freshwater. Available at: http://www.norbaf.net (accessed March 2019).

488   Kahlert, M., Albert, R.-L., Anttila, E.-L., Bengtsson, R., Bigler, C., Eskola, T., Galman, V.,

489   Gottschalk, S., Herlitz, E., Jarlman, A., Kasperoviciene, J., Kokocinski, M., Luup, H.,

490   Miettinen, J., Paunksnyte, I., Piirsoo, K., Quintana, I., Raunio, J., Sandell, B., Simola,

491   H., Sundberg, I., Vilbaste, S., Weckstrom, J . 2009. Harmonization is more important

492   than experience-results of the first Nordic-Baltic diatom intercalibration exercise 2007

493   (stream monitoring). Journal of Applied Phycology 21: 471-482.

494   Kahlert, M., Kelly, M., Albert, R.-L., Almeida, S. F. P., Bešta, T., Blanco, S., Coste, M.,

495   Denys, L., Ector, L., Fránková, M., Hlúbiková, D., Ivanov, P., Kennedy, B., Marvan, P.,

496   Mertens, A., Miettinen, J., Picinska-Faltynowicz, J., Rosebery, J., Tornés, E., Vilbaste,

497   S., Vogel, A., 2012. Identification versus counting protocols as sources of uncertainty in

498   diatom-based ecological status assessments. Hydrobiologia 695: 109–124.

499   https://doi.org/10.1007/s10750-012-1115-z.

500   Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,

501   N.E., Linder, H.P., Kessler, M., 2017a. Climatologies at high resolution for the earth's

502   land surface areas. Scientific Data 4, 170122.

503   Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,

504   N.E., Linder, H.P., Kessler, M., 2017b. Data from: Climatologies at high resolution for

505   the earth's land surface areas. Dryad Digital Repository.

506   https://doi.org/10.5061/dryad.kd1d4.

507   Kelly, M. G., Whitton, B. A., 1995. The trophic diatom index: a new index for monitoring

508   eutrophication in rivers. Journal of Applied Phycology 7: 433–444.

509   https://doi.org/10.1007/BF00003802.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson,
G. L., Solymos, P., Stevens, M. H. H., Wagner, H., 2015. Vegan: Community Ecology
Package. Available at: http://cran.r-project.org/web/packages/vegan/index.html (accessed
March 2019).

Pajunen, V., Luoto, M., Soininen, J., 2016a. Climate is an important driver for stream diatom
distributions. Global Ecology and Biogeography 25: 198–206.
https://doi.org/10.1111/geb.12399.

Pajunen, V., Luoto, M., Soininen, J., 2016b. Stream diatom assemblages as predictors of
climate. Freshwater Biology 61: 876–886. https://doi.org/10.1111/fwb.12750.

Pan, Y., Stevenson, R. J., Hill, B. H., Herhily, A. T., Collins, G. B., 1996. Using diatoms as
indicators of ecological conditions in lotic systems: a regional assessment. Journal of
North American Benthological Society 15: 481–495. https://doi.org/10.2307/1467800.

Pearman, P. B., Guisan, A., Broennimann, O., Randin, C. F., 2008. Niche dynamics in space
and time. Trends in Ecology & Evolution 23: 149–158.
https://doi.org/10.1016/j.tree.2007.11.005.

Peterson, A. T., 2011. Ecological niche conservatism: a time-structured review of evidence.
Journal of Biogeography 38: 817–827. https://doi.org/10.1111/j.1365-
2699.2010.02456.x.

Potapova, M., Charles, D. F., 2007. Diatom metrics for monitoring eutrophication in rivers of
the United States. Ecological Indicators 7: 48–70.
https://doi.org/10.1016/j.ecolind.2005.10.001.

Pulliam, H. R. (1988). Sources, sinks, and population regulation. The American Naturalist,
132, 652–661.

R Development Core Team, 2016. R: a language and environment for statistical computing. R
Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org.

535 Resh, V. H., 2008. Which group is best? Attributes of different biological assemblages used in

536   freshwater biomonitoring programs. Environmental Monitoring and Assessment 138:

537   131–138. https://doi.org/10.1007/s10661-007-9749-4.

538 Ridgeway, G. with contributions from others, 2017. gbm: Generalized Boosted Regression

539   Models. Available at: https://cran.r-project.org/package=gbm (accessed December 2018).

540 Rimet, F., Gomà, J., Cambra, J., Bertuzzi, E., Cantonati, M., Cappelletti, C., Ciutti, F.,

541   Cordonier, A., Coste, M., Delmas, F., Tison, J., Tudesque, L., Vidal, H., Ector, L., 2007.

542   Benthic diatoms in western European streams with altitudes above 800 M:

543   characterization of the main assemblages and correspondence with Ecoregions. Diatom

544   Research 22: 147–188. https://doi.org/10.1080/0269249X.2007.9705702.

545 Rott, E., Hofmann, P. G., Pall, K., Pfister, P., Pipp, E., 1997. Indikationlisten für

546   Aufwuchsalgen in Österreichischen Fliessgewässern, Teil: Sabrobielle Indikation.

547   Bundesministerium für Land- und Forstwirtschaft, Wien.

548 Sandin, L., Verdonschot, P. F. M., 2006. Stream and river typologies – major results and

549   conclusions from the STAR project. Hydrobiologia 566: 33–37.

550   https://doi.org/10.1007/s10750-006-0072-9.

551 Smol, J. P., Stoermer, E. F., 2010. The diatoms: applications for the environmental and earth

552   sciences. Cambridge University Press, xviii + 667 pp.-xviii + 67 pp.

553 Soberón, J., A. T. Peterson, 2005. Intepretation of models of fundamental ecological niches

554   and species' distributional areas. Biodiversity Informatics 2: 1–10.

555   https://doi.org/10.17161/bi.v2i0.4.

556 Soininen, J., Jamoneau, A., Rosebery, J., Leboucher, T., Wang, J., Kokocinski, M., Passy, S.

557   I., 2019. Stream diatoms exhibit weak niche conservation along global environmental

558   and climatic gradients. Ecography 42: 346–353. https://doi.org/10.1111/ecog.03828.

559 Soininen, J., Niemelä, P., 2002. Inferring the phosphorus levels of rivers from benthic

560   diatoms using weighted averaging. Archic Für Hydrobiologie 154: 1–18.

561   Soininen, J., Paavola, R., Muotka, T., 2004. Benthic diatom communities in boreal streams:

562   community structure in relation to environmental and spatial gradients. Ecography 27:

563   330–342. https://doi.org/10.1111/j.0906-7590.2004.03749.x.

564   Tan, X., Zhang, Q., Burford, M. A., Sheldon, F., Bunn, S. E., 2017. Benthic diatom based

565   indices for water quality assessment in two subtropical streams. Frontiers in

566   Microbiology 8: 601. https://doi.org/10.3389/fmicb.2017.00601.

567   Taylor, J. C., Prygien, J., Vosloo, A., de la Rey, P. A., van Rensburg, L., 2007. Can diatom-

568   based pollution indices be used for biomonitoring in South Africa? A case study of the

569   Crocodile West and Marico water management area. Hydrobiologia 592: 455–464.

570   https://doi.org/10.1007/s10750-007-0788-1.

571   Terui, A., Miyazaki, Y., Yoshioka, A., Kaifu, K., Matsuzaki, S. S., Washitani, I. 2014.

572   Asymmetric dispersal structures a riverine metapopulation of the freshwater pearl mussel

573   *Margatifera leavis*. Ecology and Evolution 4: 3004–3014.

574   Vanormelingen, P., Verleyen, E., Vyverman W., 2008. The diversity and distribution of

575   diatoms: from cosmopolitanism to narrow endemism. Biodiversity and Conservation

576   17:393–405. https://doi.org/10.1007/978-90-481-2801-3_12.

577   Werner, P., Adler, S., Dreßler, M., 2016. Effects of counting variances on water quality

578   assessments: implications from four benthic diatom samples, each counted by 40

579   diatomists. Journal of Applied Phycology 28: 2287–2297.

580   https://doi.org/10.1007/s10811-015-0760-9.

581   Wrona, F. J., Prowse, T. D., Reist, J. D., Hobbie, J. E., Lévesque, L. M. J. and Vincent, W. F.

582   2006. Climate change effects on aquatic biota, ecosystem structure and function.

583   AMBIO: A Journal of the Human Environment 35: 359–369.

584   https://doi.org/10.1579/0044-7447(2006)35[359:CCEOAB]2.0.CO;2

1   **TABLE 1.** Summary (minimum, maximum, range, median and standard deviation (Sd)) of the measured variables from 571 stream sites in
2   Sweden and 227 stream sites in Finland.
3

Variable	Unit	Sweden					Finland				
		Min	Max	Range	Median	Sd	Min	Max	Range	Median	Sd
Growing degree days (GDD)		450.9	1680.1	1229.2	1352.0	253.2	450.1	1349.1	899.0	1143.1	252.8
Frost degree days (FDD)		0	1818.1	1818.1	243.6	429.3	337.1	1811.0	1473.9	893.3	396.3
Precipitation (MAP)	mm	415.5	997.2	581.7	611.8	93.3	371.7	732.9	361.2	612.3	71.7
Water balance (WAB)	mm	102.4	784.0	681.6	255.9	98.5	126.5	441.4	314.9	274.9	48.3
Total phosphorus (TP)	$\mu g\ L^{-1}$	1.0	332.0	331.0	23.0	38.1	2.0	190.0	188.0	21.0	32.1
Conductivity	$\mu S\ cm^{-1}$	0.7	143.0	142.3	7.6	15.9	0.9	36.6	35.7	4.0	7.9
pH		3.6	9.2	5.6	7.0	0.8	4.5	8.2	3.7	6.9	0.6
Water color	$mg\ Pt\ L^{-1}$	3.5	535.0	531.5	100.0	84.2	5.0	400.0	395.0	80	71.2

4

5

6

7

1 **Table 2.** Summary of model performances shown as the model predictive ability (i.e. how
2 well the model was transformed from Sweden to Finland) (r^2). The results of the model
3 evaluations (r^2), i.e. the performances of the calibration models (in Sweden), are shown in
4 parenthesis.

	Species	**Genera**
GDD	0.05 (0.57)	0.02 (0.53)
WAB	0.00 (0.33)	0.02 (0.30)
pH	0.01 (0.65)	0.01 (0.71)
TP	0.12 (0.52)	0.10 (0.52)
Conductivity	0.13 (0.31)	0.11 (0.33)
Water color	0.04 (0.29)	0.01 (0.27)

5
6
7

8

9

10

11

12

13

14

15

16

17

18

19

20

1 **Figure legends**

2 *Figure 1.* Locations of the stream sites in Sweden (n=571, 55° – 70° N, 10° – 25° E) and

3 Finland (n=227, 60˚ – 70˚ N, 20˚ – 32˚ E) Finnish (n=227), in northern Europe.

4

5 *Figure 2.* The ordination plots of Redundancy Analysis (RDA) on the species data using the

6 climatic and environmental variables for 571 stream sites in Sweden and 227 stream sites in

7 Finland. Total phosphorus (TP), conductivity and water color were log-transformed. The

8 other abbreviations stand for growing degree days (GDD) and water balance (WAB). The

9 circles represent stream sites.

10

11 *Figure 3.* Relationships between observed and diatom species inferred values for climatic and

12 local environmental variables in the calibration models (i.e. within Sweden). Models were

13 conducted using Boosted Regression Trees (BRT) method. Each plot shows the coefficient of

14 determination (r^2) and root-mean-square error of prediction (RMSEP).

15

16 *Figure 4.* Relationships between observed and diatom species inferred values for climatic and

17 local environmental variables in Finland using the Swedish calibration models for prediction.

18 Models were conducted using Boosted Regression Trees (BRT) method. Each plot shows the

19 coefficient of determination (r^2) and root-mean-square error of prediction (RMSEP).

20

21 *Figure 5.* The ten most important species for each climatic and environmental variable in

22 Sweden and Finland. The importance of each species is shown as relative importance (%),

23 which are estimated by using Boosted Regression Trees (BRT) method.

1

2   FIG. 1

Sweden

Finland

1

2   FIG. 2

3

4

5

29

1

2   FIG. 3

1

2   FIG. 4

1

2   FIG. 5