A test for assessment of saproxylic beetle biodiversity using subsets of monitoring species

P. Sebek, T. Barnouin, Antoine Brin, Hervé Brustel, M. Dufrêne, Frédéric Gosselin, B. Meriguet, L. Micas, T. Noblecourt, O. Rose, et al.

HAL Id: hal-00794548
https://hal.science/hal-00794548

5   A test for assessment of saproxylic beetle biodiversity using subsets of "monitoring species"

6   **AUTHORS:** Pavel SEBEK[a,b], Thomas BARNOUIN[c], Antoine BRIN[d], Hervé BRUSTEL[d], Marc

7   DUFRÊNE[e,f], Frederic GOSSELIN[a], Bruno MERIGUET[g], Lilian MICAS[c], Thierry

8   NOBLECOURT[c], Olivier ROSE[c], Laurent VELLE[c], Christophe BOUGET[a,*]

9   [a] National Research Institute of Science and Technology for Environment and Agriculture.

10 (irstea), 'Forest ecosystems' Research Unit, Domaine des Barres, F-45290 Nogent-sur-Vernisson,

11 France (pav.sebek@gmail.com, frederic.gosselin@cemagref.fr, christophe.bouget@cemagref.fr)

12 [b] Department of Botany and Zoology, Masaryk University, Kotlarska 2, CZ-61137 Brno, Czech

13 Republic

14 [c] Réseau Naturaliste « Entomologie », Office National des Forêts, 2 rue Charles Péguy, F-11500

15 Quillan, France (thierry.noblecourt@onf.fr)

16 [d] Université de Toulouse, Ecole d'Ingénieurs de Purpan, INP, UMR Dynafor 1201, 75 voie du

17 T.O.E.C., BP 57611, F-31076 Toulouse Cedex 03, France (antoine.brin@purpan.fr,

18 herve.brustel@purpan.fr)

19 [e] Departement of Natural and Rural Environnement Monitoring (SPW/DGARNE/DEMNA) -

20 Avenue Maréchal Juin, 23, B-5030 Gembloux, Belgium (marc.dufrene@spw.wallonie.be)

21 [f] Liege University - Gembloux Agro Bio Tech (GxABT) - Forests, Nature, Landscape

22 Department Passage des Déportés, 2, B 5030 Gembloux

23 [g] OPIE (Office Pour les Insectes et leur Environnement), Domaine de la Minière, F-78280

24 Guyancourt, France (bruno.meriguet@insectes.org)

25 * Corresponding author: christophe.bouget@cemagref.fr

26

27

28

ABSTRACT (393)

In European forests, large scale biodiversity monitoring networks need to be implemented - networks which include components such as taxonomical groups that are at risk and that depend directly on forest stand structure. In this context, monitoring the species-rich group of saproxylic beetles is challenging. In the absence of sufficient resources to comprehensively survey a particular group, surrogates of species richness can be meaningful tools in biodiversity evaluations. In search of restricted subsets of species to use as surrogates of saproxylic beetle richness, we led a case study in Western Europe.

Beetle data were compiled from 67 biodiversity surveys and ecological studies carried out from 1999 to 2010 with standardized trapping methods in France and Belgium. This large-scale dataset contains 642 forest plots, 1521 traps and 856 species. Twenty-two simplified species subsets were identified as potential surrogates, as well as the number of genera, a higher taxonomic level, taking into account, for each surrogate, the effort required for species identification, the practical monitoring experience necessary, the species conservation potential or the frequency of species occurrence. The performance of each surrogate was analyzed based on the following parameters: overall surrogacy (correlation between subset richness and total species richness), surrogacy vs. identification cost balance, surrogacy variation over a wide range of ecological conditions (forest type, altitude, latitude and bio-geographical area) and consistency with spatial scale. Ecological representativeness and ability to monitor rare species were supplementary criteria used to assess surrogate performance.

The subsets consisting of the identifiable (or only easy-to-identify species) could easily be applied in practice and appear to be the best performing subsets, from a global point of view.

The number of genera showed good prediction at the trap level and its surrogacy did not vary across wide environmental gradients. However, the subset of easy-to-identify species and the genus number were highly sensitive to spatial scale, which limits their use in large-scale studies.

The number of rare species or the species richness of single beetle families (even the best single-family subset, the Cerambycidae) were very weak surrogates for total species richness.

Conversely, the German list of monitoring species had high surrogacy, low identification costs and was not strongly influenced by the main geographical parameters, even with our French and Belgian data.

59   In European-wide monitoring networks, such internationally validated subsets could be very

60   useful with regard to the timing and cost-efficiency of field inventories.

61

62   **Key-words (5)**

63   species richness, deadwood, biodiversity surrogate, identification costs, forest insects

64

1. INTRODUCTION

1.1. Surrogates of species richness as tools in biodiversity evaluations

In response to the global concern about an ongoing biodiversity erosion, biodiversity evaluation has become an essential part of the monitoring process to quantitatively assess publicly funded conservation programmes (selection of conservation areas, management planning), environmental monitoring schemes and biological compensation programmes (MEA, 2003). However, biodiversity evaluation is mainly based on observations such as the apparent loss of certain 'charismatic' key species or the apparent degradation of habitats in general, rather than on precisely known total species numbers (Bredemeier et al., 2007).

Species richness is one of the simplest and perhaps the most frequently used biodiversity indices because it is easy to interpret, 'sexy' enough to influence decision makers and a valuable tool for communication with the public, even though it does have certain drawbacks inherent to univariate indices (Gaston, 1996). In the absence of sufficient resources to directly carry out a comprehensive survey of a particular group of species, surrogates of species richness are used as 'shortcuts' for faunal biodiversity assessment and function as proxies to biotic distribution (Hirst, 2008; Caro, 2010).

Biodiversity surrogates may be indirect indicators of an abiotic or structural nature (environmental variables, diversity of habitats, etc.) or direct indicators closely related to the species themselves (Rodrigues and Brooks, 2007; Mandelik et al., 2012). Examples of direct surrogates of species richness include (i) the numbers of species that are frequently the focus of conservation measures (flagship, focal, indicator, keystone, and umbrella species; Favreau et al., 2006; Caro, 2010), including rare, endemic and red-listed species (Butchart et al., 2006); (ii) species subsets or combinations of these subsets which are suitable for monitoring because they are relatively easy to identify and are known to be ecologically representative (indicator species; Brin et al., 2009), or (iii) higher taxa at a reduced taxonomic resolution (genera, families, etc.), which are obviously less time-consuming to identify than species (Báldi, 2003; Balmford et al., 1996; Sebastiao and Grelle, 2009), or (iv) parataxonomic units, which are heuristically used to determine patterns in taxonomically neglected groups (Majka and Bondrup-Nielsen, 2006). Many tests relate to complementary cross-taxon surrogates, and actually test whether one subset (the surrogate taxon) is a good surrogate for another taxon (the target taxon) (Sauberer et al., 2004).

96   Attention is mostly focused on identifying a restricted subset of taxa as a surrogate of the total

97   community richness (McGeoch, 1998). Sometimes these surrogates work, but often they do not

98   (Magierowski and Johnson, 2006, Halme et al., 2009, Murphy et al., 2011). The approach is

99   suitable when the richness of only one specified group of organisms is to be studied. In this case,

100 the subsets of species may be predefined or they may be chosen randomly, for instance as a

101 restricted proportion of all the species in the taxon (Vellend et al., 2008).

102

103 *1.2. A need for surrogates of forest saproxylic biodiversity*

104 In European forests, large scale biodiversity assessments still need to be developed and

105 implemented. From the results of a test phase on intensive forest monitoring plots (Forest Biota),

106 Fischer et al. (2009) recommended (i) to include direct biodiversity components as additional

107 monitoring elements, especially for taxonomical groups that are directly dependent on trees and

108 stand structure (Seidling and Fischer, 2008), and (ii) to use the MCPFE (Ministerial Conference

109 on the Protection of Forests in Europe) criteria and indicators for sustainable forest management

110 as guidelines for the identification of additional monitoring needs.

111 Furthermore, the scientific community has been exhorting European governments for decades to

112 prioritarily use the saproxylic organisms, i.e. associated to dead or dying trees (Bouget et al.,

113 2005), in the evaluation of forest conservation status (R(88)10 and 11 recommendations from the

114 Council of Europe in 1988). Nevertheless, no practical tools are currently available to help

115 managers reach this objective. Indirect indicators such as the amount of dead wood have been

116 suggested to reflect the saproxylic beetle diversity. However, their efficient use first requires a

117 closer look at spatial scale (Lassauce et al., 2011; Schiegg, 2000; Vodka et al., 2009).

118 About a third of the European forest species need deadwood to some extent for their survival. In

119 Europe, forests have been managed for centuries (Grove, 2002), and saproxylic biodiversity,

120 which is mostly typical of and restricted to forests, is rich yet threatened due to the loss and

121 fragmentation of habitats with sufficient deadwood and veteran trees. In that, the monitoring of

122 such high saproxylic biodiversity is quite challenging.

123 Foresters and conservationists are paying more attention to beetles than to other important

124 saproxylic taxa such as fungi or Diptera; this is because more taxonomic expertise currently

125 exists in this field and surveying methods have been standardised (Majka and Bondrup-Nielsen,

126 2006). In addition, saproxylic beetles do account for a large portion of saproxylic biodiversity

127 (just behind fungi). They are widespread, numerous, species–rich and easily sampled; they

128 include representatives of many trophic guilds with a wide range of microhabitat preferences;

129 they are known to exhibit greater site specificity than vertebrates, and they often respond to

130 environmental changes more rapidly than vascular plants or vertebrates. They are therefore

131 assumed to provide valuable information on the quality and continuity of woodland habitats

132 (Grove, 2002; Nieto and Alexander, 2010; Schiegg, 2000). However, detailed taxonomic surveys

133 are often prohibitively expensive and time-consuming. Consequently, there is a need for quick

134 and easy biodiversity assessment techniques to monitor and map saproxylic beetle biodiversity.

135 Potential surrogates should be explored (Brin et al., 2009) and the cost-effectiveness of surrogates

136 must be evaluated. Cost-effectiveness assessments have already been carried out for deadwood-

137 associated polypores (Halme et al., 2009) and these assessments may provide a basis to work

138 from.

139

140 *1.3. In search of surrogates of saproxylic beetle richness: our case study in Western Europe*

141 We investigated the surrogate approach for saproxylic species richness, based on a set of beetle

142 data from temperate Western Europe. We identified and explored the performance of different

143 subsets of taxa as well as one higher taxon subset, genus richness, as surrogates of total

144 community richness. The subsets were defined mainly with respect to their potential use in future

145 exercises. We then analyzed the cost of identifying the subset and the subset sensitivity to

146 changes in environment and spatial scale. One single surrogacy result, whether positive or

147 negative, would not be a useful basis for generalization because results are likely to vary with

148 location, sampling method or forest conditions (Hess et al., 2006). To sum up, our objectives

149 were:

150 1.   to determine which simplified species subsets could be used as relevant surrogates of total

151 species richness,

152 2.   to compare subset surrogacy and identification costs,

153 3.   to check for variations in subset surrogacy over ecological and methodological gradients,

154 4.   to investigate the sensitivity of subset surrogacy to spatial scale.

155

156 **2. MATERIALS AND METHODS**

157

158   *2.1. Compiling French and Belgian datasets*

159   For our analysis, we compiled data on beetles from 67 biodiversity surveys and ecological studies

160   carried out from 1999 to 2010 by different institutes in France (Institute for Engineering in

161   Agriculture and Environment [Cemagref], National Forest Office [ONF], University of Toulouse-

162   Purpan Engineering School [EIP] and Office for Insects and their Environment [OPIE]) and the

163   Center for Research on Nature, Forest and Wood [CRNFB] in Belgium. The analysis was

164   restricted to beetle data obtained from transparent cross-vane window flight-interception traps,

165   standardised in size (Bouget et al., 2008), a widely used method for collecting saproxylic beetles.

166   The individual datasets varied greatly in their extent (number of plots, number of traps per plot,

167   number of sampling seasons). To reduce the bias due to differing study designs, we used the

168   beetle records obtained from a single trap in a single sampling season as the fundamental unit for

169   our analysis. The *trap* level records were then integrated into higher hierarchical levels, the *plot*

170   and then the *dataset* levels. In addition, we added the *region* level, to indicate the geographic

171   location of the dataset at regional scale.

172   We considered four environmental factors to describe trapping conditions: Forest type (three

173   classes: conifer, deciduous, mixed); Altitude (two classes: highland, lowland; the reference

174   altitude was 1,000 m above sea level); Latitude (two classes: north, south); and Biogeographic

175   area (four classes following the ETCBD (2006): alpine, atlantic, continental, mediterranean). We

176   also took into account the use of bait in the trap as one methodological factor (alcohol-baited,

177   unbaited). The compiled dataset derived from the beetle records in the 67 studies (66 in France,

178   one in Belgium) contained a total of 642 forest plots and 1521 traps. The distribution of the traps

179   across the environmental gradients is shown in Table 1. The location of the study plots is

180   depicted in Fig. 1 (for more detailed information about the sets used, see Appendix A).

181

182   *2.2. Data selection and standardisation*

183   The data from the different sets first had to be harmonised: nomenclature standardisation (Bouget

184   et al., 2000), removal of genus sp. The following families, often difficult to identify (Aderidae,

185   Alexiidae, Cantharidae, Carabidae, Clambidae, Corylophidae, Cryptophagidae, Dermestidae,

186   Dryopidae, Elmidae, Eucinetidae, Latridiidae, Melyridae, Mordellidae, Ptiliidae, Scirtidae,

187   Scraptiidae, Scydmaenidae, Sphaeritidae, Staphylinidae and Throscidae), were not identified at

188   the species level in all of the original datasets and were consequently removed from the compiled

189   dataset. In the Curculionidae family, only the subfamily Scolytinae was included in the analyses.

190   Four families (Ciidae, Leiodidae, Nitidulidae, Cerylonidae) were not studied in a small number

191   (from one to three) of original datasets. As these families are generally well recognized and

192   informative between saproxylic beetles, they were included in the analyses, but were analyzed

193   with a restricted compiled dataset containing only the datasets where they were studied.

194

195   *2.3. Species characterisation*

196   The 67 datasets to be analyzed contained 856 species (for a complete list see Appendix B). We

197   recorded whether or not each selected species is found on the German List of Monitoring Species

198   (Schmidl and Bussler, 2004), and on the European Red List of Saproxylic Beetles (Nieto and

199   Alexander, 2010). We defined three levels of identifiability (=ID) as levels of difficulty

200   associated to the identification of the species: 1 being the least difficult (easy to identify, e.g. by

201   picture screening), 2 requiring detailed identification keys, and 3 containing species that can be

202   identified by only a few experts (due to an insufficient literature) or requiring the preparation of

203   genitalia.

204   Species patrimoniality value (=IP) corresponded to the degree of geographic rarity in France

205   according to Bouget et al. (2010) and had four levels: (i) common and widely distributed species,

206   (ii) not abundant but widely distributed species, or only locally abundant species, (iii) not

207   abundant and only locally distributed species, (iv) very rare species (known in less than 5

208   localities or in a single 'county' in France).

209   For each species, we also included its host tree preference, feeding guild, body size class, cavity-

210   preference (Appendix B) in order to check the range of ecological groups included in each subset

211   (Does the subset reflect a wide range of forest components?). Appendix C shows the species' ID

212   and IP classes and their distribution across feeding guilds.

213   According to Mac Nally and Fleishman (2002), widespread species provide little information

214   about variance in species richness, and rare species have highly specific ecological requirements

215   that are not shared with many other species. We used the frequency of occurrence of the species

216   in the compiled dataset to create a list of 'mid frequent' species (=MidFrequent), that occurred in

217   more than 30% and less than 80% of the sets.

218   In addition, we defined the list of 'fully saproxylic' genera, that contained only saproxylic species

219   and no non-saproxylic species (see Appendix B).

220

2.4. Data subsetting

As potential surrogates for analysis, we defined 22 species subsets by crossing available species information (Tab. 2). We included different subset types. Some of them reflect the effort required for species identification (2, 10) or the monitoring experience necessary (8); others reflected the species' conservation potential, i.e. they include a high proportion of red-listed or rare species (9, 12) or the frequency of the species' occurrence (6, 11).

As for the other subsets, we isolated the 5 beetle families that occurred most often (Anobiidae, Cerambycidae, Elateridae, Curculionidae Scolytinae, Tenebrionidae). We also created subsets combining several families according to sorting error risk (18-22). Indeed, Majka and Bondrup-Nielsen (2006) found that some of these families have a low gross sorting error (Cerambycidae, Tenebrionidae, Scarabaeidae) while others have high sorting errors (Anobiidae, Nitidulidae, Curculionidae Scolytinae, Elateridae).

In the *numerous* subset, the 4 families with a high average number of individuals caught in traps were grouped; all 19 dominant families with a high average number of species caught in traps were clustered in the *domsum* subset.

Finally, we created the supra-specific surrogate of the total number of fully saproxylic genera recorded in the dataset (surrogate 23) to include a higher taxonomical level. We restricted the surrogate to fully saproxylic genera to prevent false inclusion of non-saproxylic beetles, in case the surrogate was used in practice.

2.5. Data analysis of surrogate quality

All analyses were conducted using the R software (version R 2.13, R Development Core Team, 2011).

2.5.1. Overall surrogacy

For each of the 1521 traps, we computed the total number of species found in the trap (*total species richness*), the number of species belonging to each subset and the number of fully saproxylic genera (*subset richness*). We computed R^2 between *total species richness* and *subset richness*. We used R^2 (Spearman rank correlation coefficient of determination) as a measure of explained variance.

250

251 *2.5.2. Comparison between surrogacy and costs*

252 We also evaluated costs and benefits for each species subset (excluding the genus number -

253 *sgenera*). We used the R^2 value as an indication of the benefit. Since there are no simple shortcuts

254 in collecting data in the field, and since most shortcuts occur during the identification phase, we

255 therefore attributed an identification cost value to each surrogate. We assumed that in a typical

256 community survey, the bulk of the species may be identified quickly, while a relatively small

257 number of species that are quite difficult to identify occupy a disproportionate amount of the

258 researcher's time (Colwell and Coddington, 1994). We defined 3 different cost indicators. The

259 first value (P_{sp}) was the proportion of species caught in trap, that belong to the subset, in other

260 words the proportion of species that need to be identified (a measure of lab work intensity),

261 averaged over the 1521 traps. Secondly, we took $E_{inv} = [1 -$ the proportion of species classified as

262 "easy-to-identify" from the total species list in the subset]. The lower the value, the greater the

263 proportion of easy-to-identify species in the list and the better the surrogate. Thirdly, only for the

264 best surrogates compared at the end of the process, P_{ind} was assessed as a complementary

265 measure of lab work intensity. P_{ind} was actually the proportion of individuals caught in trap,

266 whose species belong to the subset, in other words the proportion of individuals that need to be

267 identified. Nonetheless, we must keep in mind that it may be less time-consuming to identify

268 many individuals of an easy-to-identify species than only a few individuals of a difficult species.

269 Therefore, a good surrogate should be one that gives a high R^2 value with low P_{sp}, P_{ind} and E_{inv}.

270

271 *2.5.3. Variations in subset surrogacy over ecological and methodological gradients*

272 We tested the effect of environmental and methodological factors on the predictive value of our

273 surrogates. We used a mixed-effect model with a Poisson distribution, where total species

274 richness (TSR) was a dependent variable, subset richness (SSR) was an explanatory variable and

275 the environmental and methodological factors (EMF) were explanatory fixed-effect factor

276 variables (lme4 R-package). The Generalized Linear Mixed-effects model was written as follows

277 (without between-fixed effects interaction):

278 lmer(TSR~offset(log(SSR+1))+EMF+(1|region/dataset/plot).

279 The hierarchical agglomeration of trap records (plot / dataset / region) represented random

280 effects. Only one effect was tested at a time (using a Bonferroni correction related to the number

281   of tests) in order to use all records available for a given factor, even if some information was

282   missing concerning the other factors.

283

284   *2.5.4. Effect of spatial scale on predictive quality*

285   We analysed the influence of the scale on the predictive quality of the surrogates, using only

286   original datasets that contained at least 20 plots (12 datasets out of 67). The levels of plot

287   aggregation were from one to 19 plots in each dataset. At each aggregation level reflecting an

288   increasing scale, we randomly sampled the plots (without replacement) to reach the desired

289   number of plots (50 times) and then figured out the R^2 value of the surrogate. Finally, we

290   computed the mean R^2 value of the 12 datasets for each surrogate and each plot aggregation level.

291   Furthermore, we used a linear regression to test the significance of a decline in R^2 across an

292   increasing number of plots. Only the surrogates with R^2 higher than 0.85 at the trap level (as well

293   as the easy-to-identify species group) were tested for the decline.

294

295   **3. RESULTS**

296

297   *3.1. Overall surrogacy*

298   Seven of the 23 potential surrogates had R^2 greater than 0.85 (significant at p<0.001; Tab. 3).

299   They were: species from dominant families (*domsum* with R^2=0.96), mid-identifiable species

300   (excl. species difficult to identify - *identifiable* R^2=0.96), fully saproxylic genera (*sgenera*

301   R^2=0.95), common species (*common* R^2=0.90), German monitoring species (*german* R^2=0.89),

302   identifiable German monitoring species (*german identifiable* R^2=0.87) and, finally, easy-to-

303   identify, mid-identifiable and mid-frequent species (*subset4* R^2=0.86).

304   Some of the subsets explained more than 70% of the variation: the combination of the four most

305   abundant families: Anobiidae, Cerambycidae, Elateridae, Curculionidae (*numerous* R^2=0.79),

306   'mid frequent' species (*mid-frequent* R^2=0.78), easy-to-identify species (*easy-to-identify*

307   R^2=0.76), the combination of four families: Cerambycidae, Curculionidae, Mycetophagidae,

308   Nitidulidae (*combin3* R^2=0.75) and 'mid frequent' identifiable species *(identifiable mid-frequent*

309   R^2=0.73). Large species (more than 10mm in length), which are mainly identifiable species (ID1

310   or ID2), did not explain a significant proportion of the richness variation (R^2=0.62).

311   The surrogates reflecting the conservation status of species performed poorly: European Red-
312   Listed species (*iucn* $R^2=0.66$) and rare species (*rare* $R^2=0.28$). The correlation values of single
313   beetle families were non significant, the highest value being $R^2=0.50$ for Cerambycidae. Only
314   certain combinations of families showed sufficient predictive value: *combin3* ($R^2=0.75$) and
315   *numerous* ($R^2=0.79$).
316   Some of the further analyses were restricted to the 8 best performing surrogates (i.e. R² greater
317   than 0.85, and the easy-to-identify species subset).
318
319   *3.2. Comparison between surrogacy and costs*
320   R^2 prediction values over 0.85 were only achieved in the subsets where the mean proportion of
321   species to be identified (P_{sp}) was greater than 50%. From the plot comparing R^2 and P_{sp} (Fig. 2a),
322   the best ranking subsets were *german identifiable* ($R^2=0.87$; $P_{sp}=54\%$) and *subset4* ($R^2=0.86$;
323   $P_{sp}=57\%$). On the other hand, some of the surrogates provided fair prediction with a
324   comparatively low P_{sp} value, particularly *easy-to-identify* ($R^2=0.76$; $P_{sp}=29\%$). Most of the
325   subsets contained a low proportion of easy-to-identify species, from 10 to 40%. From the figure
326   2b comparing R^2 and E_{inv} (1 – the proportion of easy-to-identify species in the subset), only one
327   species subset with an R^2 value greater than 0.85, *subset4*, contained a high proportion of easy-to-
328   identify species ($R^2=0.86$; $E_{inv}=0.21$). Understandably, the *easy-to-identify* subset performed well
329   as its predictive value was relatively high ($R^2=0.76$; $E_{inv}=0$). The two subsets with the highest R^2,
330   *domsum* ($R^2=0.96$; $P_{sp}=87\%$; $E_{inv}=0.78$) and *identifiable* ($R^2=0.96$; $P_{sp}=80\%$; $E_{inv}=0.70$), turned
331   out to be very costly in both P_{sp} and E_{inv}. These subsets contained low proportions of easy-to-
332   identify species and the mean proportion of species to identify in each trap was very high.
333   Among the best scoring surrogates the *easy-to-identify* and *german identifiable* subsets
334   performed well (Tab. 5), since they contained less than half the number of individuals caught in a
335   trap to be identified on average. The other surrogates had higher P_{ind} values. The *domsum* and
336   *identifiable* subsets were very costly in P_{ind} (respectively 90% and 81% of the individuals to be
337   identified).
338
339   *3.3. Variations in subset surrogacy over ecological gradients*
340   Regarding the effect of ecological gradients, only 2 of the 8 best-performing surrogates, *easy-to-*
341   *identify* and *sgenera*, were unaffected by ecological factors. The predictive value of the other

surrogates was significantly affected by at least one of the factors (p<0.05, Tab. 3). We computed the correlation values for each factor level if the factor was significant (Tab. 4). In two of the surrogates, the R^2 value remained higher than 0.85 even if the factor effect was significant (*identifiable* - R^2 from 0.90 to 0.97; *domsum* - R^2 from 0.92 to 0.98). The predictive value of the other four surrogates fell below the level of 0.85 at least for some of the factor levels: (1) *subset4* – Forest type: conifer: R^2=0.61, Biogeographic area: continental: R^2=0.82, Biogeographic area: mediterranean: R^2=0.83; (2) *german identifiable* – Forest type: conifer: R^2=0.75, Forest type: mixed: R^2=0.84; (3) *german* – Forest type: conifer: R^2=0.77; and (4) *common* – Biogeographic area: mediterranean: R^2=0.78.

3.4. Was subset surrogacy sensitive to trap bait?

The use of bait in traps (bait factor) significantly influenced the predictive power of three quarters of the surrogates (Tab. 3). As for the 8 best performing surrogates, five of them were unaffected by bait (*easy-to-identify*, *german identifiable*, *german*, *common* and *domsum*). Three others, *subset4*, *identifiable* and *sgenera* showed lower surrogacy values in baited traps (Tab. 4). Even so, the only predictive value to fall below 0.85 was for baited traps in *subset4* (R^2=0.81).

3.5. Effect of spatial scale on surrogacy

Along the plot aggregation gradient, from the plot level to the forest scale, the decline in R^2 was significant (p<0.05) for all eight of the best-performing surrogates (Fig. 3). Except for identifiable species (*identifiable*), the R^2 in the higher plot aggregation levels fell below the 0.85 value. The R^2 of all the subsets across all aggregation levels is given in Appendix D.

4. DISCUSSION

The ideal surrogate should reflect variation in species richness well; should be easily applicable in practice; should provide consistent assessments independently of environmental gradients, methodology or spatial scale; and finally, should also sufficiently detect rare species (Colwell and Coddington, 1994; Noss, 1990). We failed to find a surrogate that would meet all the above criteria perfectly. However, attention has to be paid to individual meaning of each criterion and to perspectives of each surrogate in practical work. We summarised the global performance of the best surrogates according to 5 global criteria (Tab. 5): ecological representativeness, conservation

373   interest (i.e. the ability to monitor rare species), identification costs, surrogacy potential and

374   variability in surrogacy across environmental gradients. None of our best scoring surrogates had a

375   biased species composition regarding the host tree group, the feeding guild or the proportion of

376   cavity-dwelling species, when compared with the complete list of species.

377   The minimum level of desired surrogacy may differ according to the monitoring objectives and

378   may be adjusted to a particular use of the results. When the question is to assess if a particular

379   site is species-rich or not in comparison with other sites, a lower level of prediction could be

380   accepted. When the aim of the survey is to rank several sites according to their species richness, a

381   higher R^2 value should be required.

382

383   *4.1. Analysis of the cost-surrogacy balance*

384   The practicality of a surrogate depends on its simplicity of application. The most preferred

385   surrogates are the ones that save the most time during species identification (Magierowski and

386   Johnson, 2006) and those that can be handled by a wider spectrum of entomologists without the

387   need for experts. In our study, the most useful surrogates required the least identification time,

388   included the greatest proportion of easily identifiable species and had the fewest number of

389   species that are difficult to identify. Surrogates that explain total species richness well with a

390   limited number of species to identify per trap were deemed useful. In our study, the subsets with

391   a high proportion of easy-to-identify species showed a better cost-surrogacy balance. The number

392   of species in the subset itself is also a very important factor; subsets with many different species

393   are more difficult to use as surrogates because they require more entomological expertise.

394   Identification effort is the key consideration in biodiversity survey planning (Vellend et al.,

395   2008). From Bouget (2009), costs in working time for data collection, therefore excluding design

396   planning, data analysis and interpretation, consist in 20% field work (sampling) and 80% lab

397   work (sorting and identification). The last step itself may be divided into 20% rough sorting and

398   60% species identification. To go further, using a surrogate which saves 50% of the identification

399   costs (30% of global data collection costs) actually saves 1.5 technician hour per sample on

400   average (ca 90€in a French Institute). In our study, the best performing surrogate from a general

401   point of view was the *identifiable* subset (easy-to-identify and mid identifiable species). It had the

402   highest prediction value, showed no biases in ecological group composition, included rare species

403   and was very stable across the geographical range, the environmental gradients and the spatial

404   scale (its surrogacy potential remained very high from the trap level to the gamma aggregative

405   forest level). However, the subset contained 71% of the total number of species found in the

406   whole dataset. This makes it inherently highly probable to find many of the selected species in a

407   single trap, and it also means that the surveyor has to be able to recognise 615 species. In

408   contrast, the subset *easy-to-identify* contained only 203 species and showed quite high predictive

409   power at the trap level. Its identification costs were low since it contained only easy-to-identify

410   species and it required the identification of only 34% of the individuals and 29% of the species

411   caught in a trap on average (Fig. 2a). Using this surrogate would approximately divide the

412   identification costs by four, and consequently save 2.25 technician hour per sample on average

413   (ca 130€in a French Institute). Nonetheless, the *easy-to-identify* subset behaved poorly at higher

414   spatial scales, where its predictive ability decreased very rapidly (Fig. 3; Appendix D). The

415   global evaluation of the surrogates (Table 5) showed that the *easy-to-identify* subset, despite its

416   lower predictive value, is the most easily applicable surrogate of all. Less than 40% of the

417   individuals caught need to be identified, and yet this valuable subset is still able to reflect more

418   than 75% of the richness at the trap level on average (but only 50% of the richness at the gamma

419   scale over a forest; Fig. 3). This subset therefore deserves to be tested over wider geographical

420   and ecological gradients. Its species composition is close to the composition of another pragmatic

421   surrogate (*subset4*). Both have biased composition in favour of larger species (the proportion of

422   small species is lower than in global assemblages). The main drawback of *subset4* is its very low

423   surrogacy potential in conifer forests.

424   The higher-taxon surrogate, *sgenera* (the number of genera that contained exclusively saproxylic

425   species), also requires less identification time. Only 349 genera need to be identified, as opposed

426   to 856 species. If rare species are not an issue, *sgenera* is the second highest scoring subset (Tab.

427   5). This valuable surrogate showed good predictive power at the trap level and was not affected

428   by environmental gradients. The high correlation between genus and species richness has already

429   been reported in the literature (Báldi, 2003; Balmford et al., 1996; Gaston, 1996; Hirst, 2008)

430   though the efficiency of the higher taxa, such as families and orders, has been inconsistent (Báldi,

431   2003; Sebastiao and Grelle, 2009). In our work, to go beyond the genus level, we would have

432   been limited by the small number of exclusively saproxylic families. The scale dependence of

433   higher-taxon surrogacy (Gaston, 1996; Vanderklift et al., 1998) was well confirmed in our results

434   (Fig. 3). We therefore confirm that species-based approaches should be recommended in

435   conservation planning and monitoring when only variables at the species level are of interest

436   (ecological requirements, rarity, etc.; Bouget, 2009).

437

438   *4.2. Performance of different surrogate types*

439   We showed that surrogates based on individual beetle families were not able to predict total

440   species richness well. Conversely, Oliver and Beattie (1996) measured a significant correlation

441   between total beetle richness and the richness of three individual families. In our data, even the

442   best family surrogate, i.e. the Cerambycidae, performed poorly. Though this family is difficult to

443   sample with window trapping (Brustel, pers. comm.), it has been recommended as an indicator of

444   saproxylic beetle richness (Fayt et al., 2006); its species richness had the highest Pearson

445   correlation coefficient to the total species richness in Japanese forests (r=0.56; Ohsawa, 2010).

446   From Majka and Bondrup-Nielsen (2006), this family is less prone to low sorting, lumping and

447   splitting errors. Our best combinations (*combin3, numerous*) reached mid to high surrogacy

448   values, but remain costly (high total number of species or difficult species to identify; Tab. 5).

449   Our results suggest that biodiversity surveys should not be restricted to single beetle families but

450   should adopt a more complex approach.

451   In our data, the number of rare species was a very weak surrogate for total species richness.

452   Conversely, Müller and Goßner (2010) supported widely applying conservation practices that

453   focus on red-listed species. However, rare species by definition never appear in high numbers;

454   their prediction capability may therefore inherently be limited. Often, rare species require a

455   special sampling effort or methodology, occur unpredictably and may be under-represented in

456   collected material (Martikainen and Kouki, 2003; Martikainen and Kaila, 2004; Majka, 2007;

457   Engen et al., 2008). Biodiversity assessment based on rare species only might easily result in

458   incorrect site evaluations.

459   The 'identifiable' subset of the German monitoring species (*german identifiable*) initially

460   proposed by Schmidl and Bussler (2004) also showed relevant predictive ability. In a large-scale

461   German study, Müller and Goßner (2010) observed that their subset of 'German monitoring

462   species' mirrored the diversity pattern of the entire community, and could therefore be used in

463   state-wide monitoring studies. In our French and Belgian data, this *german identifiable* subset

464   had a high surrogacy (R²=0.87), low identification costs (46% of trapped species and 45% of

465   trapped individuals need to be identified; Tab. 5) and was not influenced by the main

466 geographical parameters but was sensitive to forest type (with a lower surrogacy in conifer than

467 in deciduous and mixed forests). In the calibration of European-wide monitoring networks (see

468 Introduction), such internationally validated subsets could be useful tools (at least in temperate

469 forests in western Europe). The definition of a standardised list of identifiable (or easy-to-

470 identify) species based on the whole checklist of European saproxylic beetle species would be a

471 profitable initiative.

472

473 *4.3. Surrogate sensitivity to environmental conditions, methodological factors and spatial scale*

474 We analysed the sensitivity of surrogate options over a wide range of ecological conditions

475 (geographic areas, forest environment, etc.). The sensitivity of biodiversity surrogates to

476 geographical location and habitat type has already been pointed out by Hess et al. (2006). Halme

477 et al. (2009) has also recently demonstrated that the subset of perennial polypores can be used as

478 a surrogate for overall polypore species richness, but that the predictive power varies in different

479 management and forest types. In our data, the surrogacy values were quite often influenced by

480 forest type (whose effect was significant on 64% of the surrogates), sometimes by the latitudinal

481 position or by the biogeographic area, and less frequently by the altitudinal position. Only half of

482 the best surrogates compared in Table 5 can be called robust across contrasting forest

483 environments. Moreover, an analysis of the magnitude of the environmental effects on surrogacy,

484 beyond their significance, could be carried out. The ecological relevance of these well-

485 performing subsets was actually improved by the wide range of feeding guilds, which reflect the

486 different ecological requirements of the species they include (see Appendix C).

487 From a methodological perspective, we demonstrated a significant effect of trap bait on

488 surrogacy for 74% of the subsets; three of the eight best surrogates were shown to be sensitive to

489 bait (Tab. 5). Indeed, some beetle species are known to be more attracted to alcohol-baited than

490 to unbaited traps (Bouget et al., 2009). Differently baited traps consequently result in different

491 species composition. Similarly, Reyers and van Jaarsveld (2000) have mentioned that assessment

492 techniques used have a strong influence on the effectiveness of biodiversity surrogates.

493 Our study indicates that some of the surrogate subsets are quite robust in terms of alpha diversity.

494 However, almost all subsets behave differently at both local and regional scales. The surrogacy of

495 only one subset (*identifiable*) was stable with increasing spatial scale. The predictive values of

496 four surrogates (*common, sgenera, subset4, easy-to-identify*) strongly decreased with increasing

497 spatial scale. In other words, subset surrogacy was reduced from the local (alpha) diversity to the

498 regional (gamma) diversity (the forest scale made up of aggregated plots). The effectiveness of

499 biodiversity indicators is known to be markedly influenced by spatial scale (Hess et al., 2006),

500 since species richness is scale dependent (Rahbek, 2005). Besides, we did not investigate the

501 contrasts between surrogates to predict the species richness of neighbouring traps, i.e. spatially-

502 associated sampling units.

503

504 *4.4. Limits and perspectives*

505 Our study dealt with surrogates of species richness (alpha and gamma diversity) and did not

506 concern species composition (beta diversity). Previous studies have demonstrated that partial

507 subsets may correlate well with total species richness, but may be less robust in reflecting species

508 composition (Magierowski and Johnson, 2006, Vellend et al., 2008). Generally, the smaller the

509 subset, the weaker the information it provides on composition patterns. From Oliver and Beattie

510 (1996), multivariate analysis of the community structure with species subsets was less powerful

511 in discriminating sites than were whole assemblages. The comparison of inter-site assemblage

512 dissimilarities with a global dataset or with simplified subsets should be carried out (Su et al.,

513 2004). Moreover, we could check that the global variations in abundance/richness of ecological

514 groups are well reflected using species subsets.

515 Some of our best performing subsets were robust in describing alpha diversity and also showed a

516 low variability in surrogacy across ecological gradients. These pragmatic surrogacy subsets may

517 be helpful in obtaining a relevant picture of total species richness in biodiversity monitoring

518 schemes. Our results do not provide an alternative to costly, but necessary, studies on species-

519 habitat relationships. However, they may provide a valuable technique to help to implement

520 effective European-wide monitoring strategies (Müller and Goßner, 2010). Designing a broad-

521 scale biodiversity monitoring program that will both address objectives and be statistically sound

522 is a significant challenge (Beggs, 2000). Even though the surrogate ability of our most robust

523 subsets remains spatially stable over our large dataset they may vary over longer time periods or

524 over wider ecological gradients. A regular reassessment of the performance of even pragmatic

525 and low-cost surrogates is required throughout any monitoring program.

526

REFERENCES:

Báldi, A., 2003. Using higher taxa as surrogates of species richness: a study based on 3700 Coleoptera, Diptera, and Acari species in Central-Hungarian reserves. Basic Appl. Ecol. 4, 589-593.

Balmford, A., Green, M.J.B., Murray, M.G., 1996. Using higher-taxon richness as a surrogate for species richness: I. Regional tests. Proc. Biol. Sci. 263, 1267-1274.

Beggs, K., 2000. Developing an Analytical Framework for Monitoring Biodiversity in Australia's Rangelands. Background paper 4. Approaches to broad scale monitoring of biological diversity – a brief review of international experience. Report from the Tropical Savannas Cooperative Research Centre.

Bouget, C. 2009. Levels of taxonomic resolution in forest entomology studies: needs, contingencies and perspectives. Mémoires de la Société Entomologique de France 8, 75 – 80 (in French)

Bouget, C., Brustel, H., Nageleisen, L.M., 2005. Nomenclature of wood-inhabiting groups in forest entomology: synthesis and semantic adjustments. C. R. Biol. 328, 936-948.

Bouget, C., Brustel, H., Brin, A., Noblecourt, T., 2008. Sampling saproxylic beetles with window flight traps: methodological insight. Rev. Ecol.-Terre Vie 63, 13-24.

Bouget, C., Brustel, H., Zagatti, P., Noblecourt, T., 2010. The French information system on saproxylic beetle ecology (FRISBEE): an ecological and taxonomical database to help with the assessment of forest conservation status. Online database available at http://frisbee.nogent.cemagref.fr/index.php/en/

Bredemeier, M., Dennis, P., Sauberer, N., Petriccione, B., Török, K., Cocciufa, C., Morabito, G., Pugnetti, A., 2007. Biodiversity assessment and change – the challenge of appropriate methods, in: Hester, R.E., Harrison, R.M. (Eds.), Biodiversity Under Threat. Issues in Environmental Science and Technology 25, Cambridge RSC Publishing, pp. 217-251.

Brin, A., Brustel, H., Jactel, H., 2009. Species variables or environmental variables as indicators of forest biodiversity: a case study using saproxylic beetles in Maritime pine plantations. Ann. For. Sci. 66, 306–316.

Butchart, S.H.M., Akcakaya, H.R., Kennedy E., Hilton-Taylor, C., 2006. Biodiversity Indicators Based on Trends in Conservation Status: Strengths of the IUCN Red List Index. Conserv. Biol.. 20, 2, 579-581

Caro, T., 2010. Conservation by Proxy: Indicator, Umbrella, Keystone, Flagship, and Other Surrogate Species. Island Press, Washington, DC

Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 345, 101-118.

Engen, S., Sæther, B.-E., Sverdrup-Thygeson, A., Grøtan, V., Ødegaard, F., 2008. Assessment of species diversity from species abundance distributions at different localities. Oikos 117, 738-748.

European Topic Centre on Biological Diversity, 2006. The indicative map of European biogeographical regions: methodology and development. Muséum National d'Histoire Naturelle, Paris.

Favreau, J.M., Drew, C.A., Hess, G.R., Rubino, M.J., Koch, F.H., Eschelbach, K.A., 2006. Recommendations for assessing the effectiveness of surrogate species approaches. Biodivers. Conserv. 15, 3949-3969.

576 Fayt, P., Dufrêne, M., Branquart, E., Hastir, P., Pontégnie, C., Henin, J.M., Versteirt, V., 2006.
577   Contrasting responses of saproxylic insects to focal habitat resources: the example of
578   longhorn beetles and hoverflies in Belgian deciduous forests. J. Insect Conserv. 10, 129-150.
579 Fischer, R., Granke, O., Chirici, G., Meyer, P., Seidling, W., Stofer, S., Corona, P., Marchetti,
580   M., Travaglini, D., 2009. Background, main results and conclusions form a test phase for
581   biodiversity assessments on intensive forest monitoring plots in Europe. iForest-
582   Biogeosciences and Forestry 2, 58–65.
583 Gaston, K.J. (Ed.), 1996. Biodiversity: Ecology of Numbers and Difference. Blackwell Science,
584   London.
585 Grove, S.J., 2002. Saproxylic insect ecology and the sustainable management of forests. Annu.
586   Rev. Ecol. Syst. 33, 1-23.
587 Halme, P., Kotiaho, J.S., Ylisirniö, A.L., Hottola, J., Junninen, K., Kouki, J., Lindgren, M.,
588   Mönkkönen, M., Penttilä, R., Renvall, P., Siitonen, J., Similä, M., 2009. Perennial polypores
589   as indicators of annual and red-listed polypores. Ecol. Indic. 9, 256-266.
590 Hess, G.R., Bartel, R.A., Leidner, A.K., Rosenfeld, K.M., Rubino, M.J., Snider, S.B., Ricketts,
591   T.H., 2006. Effectiveness of biodiversity indicators varies with extent, grain, and region. Biol.
592   Conserv. 132, 448-457.
593 Hirst, A.J., 2008. Surrogate measures for assessing cryptic faunal biodiversity on macro-algal
594   dominated subtidal reefs. Biol. Conserv. 141, 211-220.
595 Lassauce, A., Paillet, Y., Jactel, H., Bouget, C., 2011. Deadwood as a surrogate for forest
596   biodiversity: Meta-analysis of correlations between deadwood volume and species richness of
597   saproxylic organisms. Ecol. Indic. 11, 1027-1039.
598 Mac Nally, R., Fleishman, E., 2002. Using "indicator" species to model species richness: Model
599   development and predictions. Ecol. Appl. 12, 79-92.
600 Magierowski, R.H., Johnson, C.R., 2006. Robustness of surrogates of biodiversity in marine
601   benthic communities. Ecol. Appl. 16, 2264-2275.
602 Majka, C.G., 2007. The Eucnemidae (Coleoptera) of the Maritime provinces of Canada: new
603   records, observations on composition and zoogeography, and comments on the rarity of
604   saproxylic beetles. Zootaxa 1636, 33–46.
605 Majka, C.G., Bondrup-Nielsen, S., 2006. Parataxonomy: a test case using beetles. Animal
606   Biodivers. Conserv. 29, 149–156.
607 Mandelik, Y., Dayan, T., Chikatunov, V., Kravchenko, V., 2012. The relative performance of
608   taxonomic vs. environmental indicators for local biodiversity assessment: A comparative
609   study. Ecol. Indic. 15, 1, 171-180
610 Martikainen, P., Kaila, L., 2004. Sampling saproxylic beetles: lessons from a 10-year monitoring
611   study. Biol. Conserv. 120, 171-181.
612 Martikainen, P., Kouki, J., 2003. Sampling the rarest: threatened beetles in boreal forest
613   biodiveristy inventories. Biodivers. Conserv. 12, 1815-1831.
614 McGeoch, M.A., 1998. The selection, testing and application of terrestrial insects as
615   bioindicators. Biol. Rev. Camb. Philos. Soc. 73, 181-201.
616 M.E.A.(Millennium Ecosystem Assessment), 2003. Ecosystems and Human Wellbeing. A
617   Framework for Assessment. Island Press, Washington DC. 245 p.
618 Müller, J., Goßner, M.M., 2010. Three-dimensional partitioning of diversity informs state-wide
619   strategies for the conservation of saproxylic beetles. Biol. Conserv. 143, 625-633.
620 Murphy, D., Weiland, P.S., Cummins, K.W., 2011. A Critical Assessment of the Use of
621   Surrogate Species in Conservation Planning in the Sacramento-San Joaquin Delta, California
622   (USA). Conserv. Biol., 25, 5, 873-878

623   Nieto, A., Alexander, K.N.A., 2010. European Red List of Saproxylic Beetles. Publication Office
624   of the European Union, Luxembourg.
625   Noss, R.F., 1990. Indicators for monitoring biodiversity: A hierarchical approach. Conserv. Biol.
626   4, 355-364.
627   Oliver, I., Beattie, A.J., 1996. Designing a cost-effective invertebrate survey: a test of methods
628   for rapid assessment of biodiversity. Ecol. Appl. 6, 594–607.
629   Ohsawa, M., 2010. Beetle families as indicators of Coleopteran diversity in forests: a study using
630   Malaise traps in the central mountainous region of Japan. J. Insect Conserv. 14, 479-484
631   Rahbek, C., 2005. The role of spatial scale and the perception of large-scale species-richness
632   patterns. Ecol. Lett. 8, 224-239.
633   Reyers, B., van Jaarsveld, A.S., 2000. Assessment techniques for biodiversity surrogates. S. Afr.
634   J. Sci. 96, 406–408.
635   Rodrigues, A.S.L., Brooks, T.M., 2007. Shortcuts for biodiversity conservation planning: The
636   efectiveness of surrogates. Annu. Rev. Ecol. Evol. Syst. 38, 713-737.
637   Sauberer, N., Zulka, K.P., Abensperg-Traun, M., Berg, H.M., Bieringer, G., Milasowszky, N.,
638   Moser, D., Plutzar, C., Pollheimer, M., Storch, C., Tröstl, R., Zechmeister, H., Grabherr, G.,
639   2004. Surrogate taxa for biodiversity in agricultural landscapes of eastern Austria. Biol.
640   Conserv. 117, 181–190.
641   Schiegg, K., 2000. Are there saproxylic beetle species characteristic of high dead wood
642   connectivity? Ecography 23, 579-587.
643   Schmidl, J., Bussler, H., 2004. Ökologische Gilden xylobionter Käfer Deutschlands und ihr
644   Einsatz in der landschaftsökologischen Praxis – ein Bearbeitungsstandard. Naturschutz und
645   Landschaftsplanung 36, 202-218.
646   Sebastiao, H., Grelle, C.E.V., 2009. Taxon surrogates among Amazonian mammals: can total
647   species richness be predicted by single orders? Ecol. Indic. 9, 160–166
648   Seidling, W., Fischer, R., 2008. Deviances from expected Ellenberg indicator values for nitrogen
649   are related to N throughfall deposition in forests. Ecol. Indic. 8, 639-646.
650   Su, J.C., Debinski, D.M., Jakubauskas, M.E., Kindscher, K., 2004. Beyond species richness:
651   community similarity as a measure of cross-taxon congruence for coarse-filter conservation.
652   Conserv. Biol. 18, 167–173.
653   Vanderklift, M.A., Ward, T.J., Phillips, J.C., 1998. Use of assemblages derived from different
654   taxonomic levels to select areas for conserving marine biodiversity. Biol. Conserv. 86, 307-
655   315.
656   Vellend, M., Lilley, P.L., Starzomski, B.M., 2008. Using subsets of species in biodiversity
657   surveys. J. Appl. Ecol. 45, 161-169.
658   Vodka, S., Konvicka, M., Cizek, L., 2009. Habitat preferences of oak-feeding xylophagous
659   beetles in a temperate woodland: implications for forest history and management. J. Insect
660   Conserv. 13, 553-562.
661

662   **Table legends**
663
664   **Table 1**: Trap numbers in categories representing environmental gradients; based on
665   Biogeographic area, Forest type and Altitude.
666
667   **Table 2:** Surrogate description. IP 1-2 = number of species with patrimoniality value 1 and 2, IP
668   3-4 = number of species with patrimoniality value 3 and 4, MidFrequent = number of species
669   present on the list of 'mid frequent' species (see section 2.3. of the text), ID-1 = number of easy-
670   to-identify species, ID-2 = number of mid-identifiable species, * = number of genera.
671
672   **Table 3:** Surrogacy values and sensitivity to environmental and methodological factors of the 42
673   subsets. Results of the Spearman correlation between number of species in subsets and total
674   species richness (R^2 is displayed; all the correlations were significant at 0.001 level) and the
675   effect of environmental factors on predictions (level of significance for factor effects: ** =
676   p<0.01; *** = p<0.001; dash = not significant).
677
678   **Table 4:** Correlation influenced by factors. The numbers in the table give the R^2 values for each
679   level of ecological and methodological factor variable. The values are displayed only where the
680   effect of the factor was significant. Only the surrogates with R^2 greater than 0.85 (as well as the
681   easy-to-identify subset) were included in the table. The values below 0.85 are in bold.
682
683   **Table 5:** Global surrogate evaluation according to the following criteria: (i) ecological
684   representativeness, (ii) conservation interest, (iii) identification costs, (iv) surrogacy potential and
685   (v) variability in surrogacy across environmental gradients
686   The fields are as follows: groups by preferred host tree, feeding guilds
687   (Saproxylo.=saproxylophagous, Sec. xylo.=secondary wood decayer), large species (>10mm in
688   body size), cavity-dwelling species, common species (IP=1 or 2), rare species (IP=3 or 4).
689   Numbers indicate the proportion (%) of the species in the subset. Variability in surrogacy
690   concerns the effects of forest type, geographical range (latitude, biogeographical region, altitude),
691   alcohol bait, increasing spatial scale (plot aggregation up to the whole forest level) on R^2 values.
692   Indices indicate the performance of the surrogate for each of the 5 criteria mentioned above: 2 =
693   good (well-performing surrogate in the criterion), 1 = average, 0 = bad (badly-performing
694   surrogate in the criterion). P_{ind} (= proportion of individuals to be identified in the species subset
695   from all individuals caught in the trap); P_{sp} (= proportion of species to be identified in the subset
696   from all species caught in the trap); E_{inv} = [1 - the proportion of easy-to-identify species from the
697   total species list in the subset], R^2 = correlation value between subset and total species richness,
698   Sum = Cumulative index (the higher the value, the better), * = estimated value for the higher-
699   taxon surrogate.
700

Figure legends

Fig. 1: Distribution of the plots that were part of the study. Dotted line shows the reference border used for segregation between latitudinal levels (north and south), the altitudinal affiliation of each plot is displayed in full black (lowland) or grey-filled points (highland). The separation into biogeographic areas is depicted by polygons. *The plots (point numbers) were part of the following datasets: 1 – Assise; 2 – Ballons-Comtois; 3 – Chaumes; 4 – Gerardmer; 5 – Kertoff; 6 – St-Maurice; 7 – Tourbière-Charmes; 8 – Ventron; 9 – Brie; 10 – Courneuve; 11, 41, 42 – East-France; 12 – Fleury-Biere; 13 – Fontainebleau; 14 – Seine-et Marne (Fontaine-Port); 15 – Seine-et-Marne (Fontainebleau-Opie); 16 – Hauts-de-Seine; 17 – Seine-et-Marne (Larchant-Marais); 18 – Lay; 19 – Seine-et-Marne (Livry); 20 – Seine-et-Marne (Maincy); 21 – Maussoin, 22 – Seine-et-Marne (Noyen-sur-Seine); 23 – Orléans; 24 – Orléans-Regix; 25 – Rambouillet; 26 – Sausset; 27 – Thiercelieux; 28 – Seine-et-Marne (Valence); 29 – Attigny; 30 – Auberive; 31 – Bannes; 32 – Belgium-Wallonie; 33 – Bois-de-Champ; 34 – Bresse; 35 – Chalmessin; 36 – Chatillon; 37 - Chaux-Regix; 38 – Colettes; 39 – Combe-Lavaux; 40 – Darney; 43 – Gehant; 44 – Haute-Meurthe; 45 – Messarges; 46 – Mont-Dieu; 47 – Mortagne; 48 – Relanges; 49 – RNVA; 50 – Ternes; 51 - Trois-Fontaines; 52 - Troncais-Cem; 53 – Troncais (ONF); 54 – Troncais (Velle); 55 - Val-de-Senones; 56 – Vauhalaise; 57 – Canopee; 58 - Haute-Savoie; 59 – Jujols; 60 – Mantet; 61 – Mercantour; 62 – Natura 2000; 63 – Orlu; 64 – Tete-Alpes; 65 – Auvergne; 66 – Lozere; 67 – Bléones; 68 – Chamatte; 69 – Luberon; 70 – Tartonne; 71 – Caylus; 72 – Landes; 73 – Cadarache; 74 – Maures (ONF); 75 – Maures (Purpan)*

Fig. 2: Comparison between surrogacy and cost parameters of the surrogates. (a) between R^2 and P_{sp} (= proportion of species to be identified in the subset from all species caught in the trap); (b) between R^2 and E_{inv} = [1 - the proportion of easy-to-identify species from the total species list in the subset]. Note that only subsets with R^2 greater than 0.5 are displayed. The dashed line shows the value of R^2=0.85. *Point numbers correspond to these subsets: 1 - easy mid-frequent; 2 - easy-to-identify; 3 - identifiable mid-frequent; 4 - subset4; 5 – identifiable; 6 - mid-frequent; 7 - german identifiable; 8 - german; 9 – iucn; 10 - larger10; 11 – common; 18 - combin1; 19 - combin2; 20 - combin3; 21 – numerous; 22 – domsum*

Fig. 3: Scale effect. Prediction ability of surrogates (mean R^2 value of 12 datasets) across an increasing number of aggregated plots. Only surrogates with R^2 greater than 0.85 at the trap level are displayed. The dashed line shows the value of R^2=0.85.

736 **Table 1**

737

Biogeographic Area	Forest type	Altitude		Total trap no
		highland	lowland	
Alpine	conifer	35	-	35
	deciduous	52	-	52
	mixed	64	-	64
Atlantic	conifer	-	99	99
	deciduous	-	522	522
	mixed	-	34	34
Continental	conifer	2	-	2
	deciduous	4	461	465
	mixed	92	36	128
Mediterranean	conifer	1	-	1
	deciduous	58	28	86
	mixed	33	-	33
Total		341	1180	1521

Table 2

Surrogate	name	species no	IP 1-2	IP 3-4	MidFrequent	ID-1	ID-2	Description
1	easy mid-frequent	32	31	1	32	32	0	easy-to-identify mid frequent species (ID=1 x MidFrequent)
2	easy-to-identify	203	151	52	32	203	0	easy-to-identify species (ID=1)
3	identifiable mid-frequent	84	80	4	84	32	52	easy-to-identify mid frequent species and mid identifiable mid frequent species (ID=1*MidFrequent + ID=2*MidFrequent)
4	subset4	255	200	55	84	203	52	easy-to-identify species and mid identifiable mid frequent species (ID=1 + ID=2*MidFrequent)
5	identifiable	615	444	171	84	203	412	easy-to-identify species and mid identifiable species (ID=1 + ID=2)
6	mid-frequent	102	98	4	102	32	52	mid frequent species (MidFrequent)
7	german identifiable	394	276	118	65	150	244	easy-to-identify and mid identifiable species present on the German list of monitoring species (Schmidl & Bussler 2004)
8	german	499	359	140	74	150	244	all species present on German list of monitoring species
9	iucn	173	111	62	27	63	99	species present on the list of European red-listed species (Nieto & Alexander 2010)
10	larger10	193	147	46	24	99	79	species with body size greater than or equal to 10 mm
11	common	256	256	0	63	65	118	common species (IP=1)
12	rare	217	0	217	4	52	119	rare species (IP=3 or IP=4)
13	Anobiidae	72	46	26	5	0	21	species family Anobiidae
14	Cerambycidae	149	111	38	16	69	77	species family Cerambycidae
15	Curculionidae	95	89	6	13	26	44	species in family Curculionidae
16	Elateridae	49	25	24	8	12	26	species family Elateridae
17	Tenebrionidae	55	46	9	3	9	35	species family Tenebrionidae
18	combin1	268	197	71	29	100	143	Cerambycidae, Elateridae, Scarabaeidae, Tenebrionidae (generally well known saproxylic species with different habitat preferences)
19	combin2	355	254	101	37	101	177	Cerambycidae, Elateridae, Scarabaeidae, Tenebrionidae, Mycetophagidae, Anobiidae (*combin1* extended)
20	combin3	296	243	53	36	97	147	Cerambycidae, Curculionidae, Mycetophagidae, Nitidulidae (families performing well in preliminary analyses)
21	numerous	365	271	94	88	107	168	Anobiidae, Cerambycidae, Elateridae, Curculionidae (families with a high average number of individuals caught in traps)
22	domsum	653	482	171	42	149	320	All 19 dominant families
23	sgenera	349*	-	-	-	-	-	number of fully saproxylic genera

Table 3

Surrogate	R²	Ecological effects				Methodological effect
		Forest type	Altitude	Latitude	Biogeographic area	Bait
easy mid-frequent	0.59	***	**	***	**	***
easy-to-identify	0.76	-	-	-	-	-
identifiable mid-frequent	0.73	***	**	***	-	**
subset4	**0.86**	***	-	-	-	**
identifiable	**0.96**		-	-	-	**
mid-frequent	0.78	***	-	***	**	**
german identifiable	**0.87**	***	-	-	-	-
german	**0.89**	***	-	-	-	-
iucn	0.66	***	**	-	**	**
larger10	0.62	-	-	-	-	-
common	**0.90**	-	-	***	***	-
rare	0.28	-	-	-	-	***
Anobiidae	0.16	-	-	-	**	***
Cerambycidae	0.50	-	-	-	-	-
Curculionidae	0.30	***	-	-	-	***
Elateridae	0.26	**	-	**	-	***
Tenebrionidae	0.16	-	-	-	-	***
combin1	0.60		-	**	-	-
combin2	0.69	**	-	**	-	-
combin3	0.75	***	-	-	-	**
numerous	0.79	***	-	-	**	-
domsum	**0.96**	-	-	-	***	-
sgenera	**0.95**	-	-	-	-	***

Table 4

Factor	Level	Surrogate							
		subset4	easy-to-identify	identifiable	german identifiable	german	common	domsum	sgenera
Forest type	conifer	**0.61**	-	-	**0.75**	**0.77**	-	-	-
	deciduous	0.89	-	-	0.91	0.92	-	-	-
	mixed	0.89	-	-	**0.84**	0.87	-	-	-
Altitude	highland	-	-	-	-	-	-	-	-
	lowland	-	-	-	-	-	-	-	-
Latitude	north	-	-	-	-	-	0.89	-	-
	south	-	-	-	-	-	0.92	-	-
Biogeographic area	alpine	-	-	-	-	-	0.93	0.92	-
	atlantic	-	-	-	-	-	0.90	0.98	-
	continental	**-**	-	-	-	-	0.92	0.93	-
	mediterranean	**-**	-	-	-	-	**0.78**	0.94	-
Bait	baited	**0.81**	-	0.91	-	-	-	-	0.90
	not baited	0.87	-	0.97	-	-	-	-	0.97

Table 5

			identifiable	easy-to-identify	subset4	domsum	sgenera	german	german identifiable	common	Total
(i) Ecological representativeness	Host tree groups	Conifer	14 %	14 %	12 %	17 %		13 %	12 %	19 %	16 %
		Deciduous	56 %	60 %	60 %	51 %		62 %	61 %	50 %	53 %
	Feeding guild	Mycophagous	17 %	12 %	15 %	14 %		19 %	19 %	21 %	17 %
		Predator	18 %	19 %	19 %	17 %		23 %	22 %	11 %	18 %
		Saproxylo.	15 %	13 %	12 %	13 %		11 %	11 %	13 %	13 %
		Sec. xylo.	44 %	52 %	50 %	44 %		45 %	45 %	45 %	42 %
		Large	29 %	49 %	42 %	25 %		29 %	34 %	18 %	23 %
		Cavity-dwelling	12 %	12 %	11 %	10 %		15 %	15 %	8 %	11 %
	Index		2	1	1	2	2	1	1	2	
(ii) Conservation interest	IP	Common	72 %	74 %	78 %	74 %		72 %	70 %	100 %	75 %
		Rare	28 %	26 %	22 %	26 %		28 %	30 %	0 %	25 %
	Index		2	2	2	2	0	2	2	0	
(iii) Identification costs		E_{inv}	0.70	0.00	0.21	0.78		0.79	0.64	0.80	0.24
		P_{sp}	80 %	29 %	57 %	87 %		66 %	54 %	62 %	
		P_{ind}	81%	34%	59%	90%		55%	45%	69%	
		Total no of species	615	203	255	653	349	499	394	256	856
	Index		1	2	2	0	1*	1	1	1	
(iv) Surrogacy potential		R^2	0.96	0.76	0.86	0.96	0.95	0.89	0.87	0.90	
	Index		2	1	2	2	2	2	2	2	
(v) Variability in surrogacy		with forest type & geographical range	Stable	Stable	Lower R² in conifer forests	Lower R² in alpine area	Stable	Lower R² in conifer forests	Lower R² in conifer/mixed forests	Lower R² in Mediterranean area	
		with trap baiting	Lower R² in baited trap data	Stable	Lower R² in baited trap data	Stable	Lower R² in baited trap data	Stable	Stable	Stable	
		with spatial scale	stable	highly unstable	highly unstable	Unstable	highly unstable	Unstable	Unstable	highly unstable	
	Index		2	1	0	1	1	0	0	0	
Sum (max=10)			**9**	**7**	**7**	**7**	**6**	**6**	**6**	**5**	