

1 Predicting habitat affinities of plant species using commonly measured
2 functional traits

3 Bill Shiple and Michael Belluau, Département de biologie, Université de Sherbrooke, Sherbrooke (Qc)
4 J1K J1K 2R1 Canada Bill.Shipley@USherbrooke.ca; Michael.Belluau@USherbrooke.ca

5 Dr. Ingolf Kühn, Dept. Community Ecology, Helmholtz Centre for Environmental Research GmbH - UFZ,
6 Theodor-Lieser-Str. 4, 06120 Halle , Germany. ingolf.kuehn@ufz.de

7
8 [Nadejda A. Soudzilovskaia](mailto:Nadejda.A.Soudzilovskaia@cml.leidenuniv.nl), Conservation Biology Department, Institute of Environmental Sciences, CML,
9 Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands n.a.soudzilovskaia@cml.leidenuniv.nl

10 Michael Bahn, Institute of Ecology, University of Innsbruck, Sternwartestr. 15, 6020 Innsbruck, Austria,
11 Michael.Bahn@uibk.ac.at

12 Josep Penuelas: CSIC, Global Ecology Unit CREAM-CSIC-UAB, Cerdanyola del Vallès (Catalonia), Spain;
13 CREAM, Cerdanyola del Vallès, Barcelona, Catalonia, Spain, josep.penuelas@uab.cat

14 Jens Kattge, Max Planck Institute for Biogeochemistry, Hans Knöll Str. 10, 07745 Jena, Germany AND
15 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103
16 Leipzig, Germany; jkattge@bgc-jena.mpg.de

17 Lawren Sack, Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 621
18 Charles E. Young Drive South, Los Angeles, California, USA 90025 lawrensack@ucla.edu

19 Jeannine Cavender-Bares, Department of Ecology, Evolution and Behavior, University of Minnesota,
20 Saint Paul MN, USA; cavender@umn.edu

21 Wim A. Ozinga, Team Vegetation, Forest and Landscape Ecology, Alterra, Wageningen UR, PO Box 47
22 NL-6700 AA Wageningen , The Netherlands wim.ozinga@wur.nl

23 Benjamin Blonder, Environmental Change Institute, School of Geography and the Environment,
24 University of Oxford, South Parks Road, Oxford OX1 3QY, United Kingdom. bblonder@gmail.com.

25 Peter M. van Bodegom, Institute of Environmental Sciences, CML, Leiden University, Einsteinweg 2,
26 2333 CC Leiden, The Netherlands p.m.van.bodegom@cml.leidenuniv.nl

27 Peter Manning, Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre
28 (BiK-F), Senckenberganlage 25, D-60325 Frankfurt, Germany, peter.manning@ips.unibe.ch

29 Thomas Hickler, Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25,
30 D-60325 Frankfurt am Main, Germany & Department of Physical Geography, Geosciences, Goethe-
31 University Frankfurt am Main, Germany; thomas.hickler@senckenberg.de

32 Enio Sosinski, Embrapa Clima Temperado, Pelotas, RS, Brazil, 96010-971 enio.sosinski@embrapa.br

33

34 Valério De Patta Pillar, Department of Ecology, Universidade Federal do Rio Grande do Sul

Shiple et al.

35 Porto Alegre, RS, 91540-000, BRAZIL vpillar@ufrgs.br

36 Vladimir Onipchenko, Geobotany department, Moscow State University vonipchenko@mail.ru

37

38 Words: 6845

39 Figures: 3

40 Tables: 1

41

This is the accepted version of the following article: Shiple, B. et al. "Predicting habitat affinities of plant species using commonly measured functional traits" in Journal of Vegetation Science, vol. 28, issue 5 (Sep. 2017), p. 1082-1095, which has been published in final form at DOI 10.1111/jvs.12554. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

42 **Abstract**

43 **Questions:** : Heinz Ellenberg classically defined “indicator” scores for species representing their typical
44 positions along gradients of key environmental variables, and these have proven very useful for
45 designating ecological distributions. We tested a key tenent of trait-based ecology, i.e., the ability to
46 predict ecological preferences from species’ traits. More specifically, can we predict Ellenberg indicator
47 scores for soil nutrients, soil moisture and irradiance from four well-studied traits: leaf area, leaf dry
48 matter content, specific leaf area and seed mass? Can we use such relationships to estimate Ellenberg
49 scores for species never classified by Ellenberg?

50 **Location:** Global

51 **Methods:** Cumulative link models were developed to predict Ellenberg nutrients, irradiance and
52 moisture values from Ln-transformed trait values using 922, 981 and 988 species respectively. We then
53 independently tested these prediction equations using the trait values of 423 and 421 new species that
54 occurred elsewhere in Europe, North America and Morocco and whose habitat affinities we could classify
55 from independent sources as 3-level ordinal ranks related to soil moisture and irradiance. The traits
56 were specific leaf area, leaf dry matter content, leaf area and seed mass.

57 **Results:** The four functional traits predicted the Ellenberg indicator scores of site fertility, light and
58 moisture with average error rates of < 2 Ellenberg ranks out of 9. We then used the trait values of 423
59 and 421 species respectively that occurred (mostly) outside of Germany but whose habitat affinities we
60 could classify as 3-level ordinal ranks related to soil moisture and irradiance. The predicted positions of
61 the new species, given the equations derived from the Ellenberg indices, agreed well with their
62 independent habitat classifications, although our equation for Ellenberg irradiance levels performed
63 poorly on the lower ranks.

Shipley et al.

64 **Conclusions:** These prediction equations, and their eventual extensions, could be used to provide
65 approximate descriptions of habitat affinities of large numbers of species worldwide.

66 Key words: environmental gradients, habitat affinities, soil moisture, soil nutrients, habitat fertility,
67 shade, wetlands, understory plants, SLA, LDMC, leaf size, seed size.

68

69

70 **Introduction**

71 Plant ecology, emerging from plant geography, has existed as an academic discipline for more
72 than a century (Warming et al. 1909). A major recurring theme in this discipline is that plant species are
73 differentially distributed along important abiotic gradients with respect to the most common
74 environmental variables that control plant growth and survival. Given this history, one might think that
75 basic information concerning the range and modal values of most plant species along such abiotic
76 gradients (i.e. their habitat “affinities”) would be generally available. In fact, such basic ecological
77 information is often missing with the exception of species like aquatics, desert species, or understory
78 herbs that are restricted to clearly differentiated habitats.

79 One exception is the German flora. Heinz Ellenberg (1988, 1978, 1992), working with this flora,
80 defined a series of ordinal scales for seven major environmental variables (irradiance level; soil
81 moisture, pH, nitrogen availability, soil salinity; average yearly temperature and continentality). He then
82 assigned a single “indicator” score for each species with respect to each environmental value identifying
83 position along the environmental gradient at which the species was most common. These indicator
84 scores can be viewed collectively as very approximate descriptions of the mode of the realized
85 Hutchinsonian niche (Hutchinson 1957) of each species. The scales are ordered and go from one (lowest
86 level) to nine (highest level) except for moisture which goes to 12. We use three Ellenberg indicator
87 scores that are directly linked to resource capture: “nitrogen” (better interpreted as a description of site
88 productivity or soil nutrient levels), “irradiance” and “moisture”. Ellenberg values have proven very
89 popular since Ellenberg’s (1988, 1978) original references have been cited over 7000 times in the
90 ecological literature.

91 These ordinal scales, and the indicator value assigned to a species for a given scale, were not
92 based on direct physical measurements of the environment but rather on Ellenberg’s own field

Shiple et al.

93 observations with input from his collaborators. Ellenberg indicator scores, being based on expert
94 opinion rather than direct physical measurement of the abiotic variables, require the sort of extensive
95 and exhaustive knowledge of a local flora that takes a lifetime to obtain, and that few people possess.
96 Their subjective nature also makes them very difficult to generalize to new geographical locations in a
97 way that is comparable (Valladares et al. 2008). Strictly speaking, since these are expert *opinions* rather
98 than objective *measurements*, the only people who could to truly extend such indicator scores to new
99 locations would be Heinz Ellenberg (who died in 1997), or someone who trained with him. Similar, but
100 not identical, schemes have been produced for other European countries such as Austria (Karrer 1992),
101 Switzerland (Landolt 1977) Italy (Guarino et al. 2012, Pignatti 2005) and Poland (Zarzycki et al. 2002) .

102 Despite the subjective nature of Ellenberg indicator scores, a number of independent
103 publications have shown that average Ellenberg numbers, based on weighted or unweighted vegetation
104 relevés, do correlate well with more explicit environmental measurements (Bartelheimer et al. 2014,
105 Carpenter et al. 2014, Douma et al. 2012, Ertsen et al. 1998, Franzaring et al. 2007). Mean Ellenberg
106 indicator scores of plots in a large vegetation dataset representing all terrestrial habitats across the
107 Netherlands were also highly correlated to the position of plots across the main gradients in species
108 composition as revealed by ordination models (Ozinga et al. 2005). Although Ellenberg indicator scores
109 are only approximate and non-metric descriptions of the mode of a species' distribution along such
110 environmental gradients, they might better integrate the effect of temporal and spatial variation of such
111 environmental conditions on plant performance in ways that more explicit but short-term
112 environmental measurements cannot. Some of the environmental properties referenced by Ellenberg
113 indicator scores can be measured using well-established methods and are sufficiently stable in time and
114 well-defined (soil pH or salt content) that they can be directly measured using samples taken at a single
115 point in time. Habitat affinities based on such direct measurements are rare in practice because one
116 must take them over large spatial extents in order to properly quantify the distribution of a species

117 along such gradients and to identify the values at which a species reaches maximum abundance. Other
118 variables, such as soil moisture, irradiance level or temperature, can be measured using well-established
119 methods. However, they are so temporally and spatially variable that one would require both spatially
120 extensive sampling and long-term monitoring in order to quantify them in a way that is meaningful to
121 plant performance; such studies are even rarer. Finally, even if the above practical problems can be
122 overcome, there does not even exist a generally-accepted method of quantifying an important soil
123 property like soil nutrient availability (or “fertility”), which is determined by a complicated interaction of
124 both a number of soil properties (Fujita et al. 2013) and plant and fungal properties that enable uptake
125 of different limiting nutrients. Such difficulties explain, in large part, why we do not yet possess even
126 basic information concerning the habitat affinities of most plant species and why we are unlikely to fill
127 this knowledge gap in the near future. If this is true, then can we generalize this expert knowledge to
128 new geographical areas?

129 One method, developed by Hill et al (1999) for the British flora, uses co-occurrence data of
130 vegetation, in conjunction with the Ellenberg indicator scores of the subset of species for which this
131 information is available, to generalize to new species and to adjust the values of other species as
132 needed. Although this method worked well for the British flora, which contains many species in
133 common with the German flora, it would not work when extrapolating to geographical areas with few or
134 no species for which Ellenberg numbers are available. Another method (Klaus et al. 2012) of
135 generalizing Ellenberg numbers to new locations was based on calibration curves linking Near Infrared
136 Spectrometric signatures (NIRS) of vegetation samples in Germany with averaged Ellenberg values of
137 these samples. In this paper we evaluate a trait-based method for predicting Ellenberg indicator scores
138 and evaluate its potential to predict habitat affinities of new species around the world for which
139 Ellenberg indicator scores do not exist.

140 A functional trait is any phenotypic attribute, measurable on an individual plant, that affects its
141 growth, survival or reproductive success and thus its fitness in a given environmental context (Violle et
142 al. 2007). Extrapolating to all individuals of a given species in a local area, such functional traits affect
143 the demographic performance of the species and so a basic postulate of trait-based ecology is that a
144 species' demographic performance along major environmental gradients is partly determined by the
145 values of at least some of its functional traits at some stage in its ontogeny (Grime 1979). Ideally, when
146 estimating a species' habitat from its traits, one would choose traits that are simultaneously the most
147 important mechanistic determinants of plant fitness along the chosen environmental gradients, are
148 widely measured for many plant species, and are easily measurable for new species. In practice these
149 criteria are often strongly conflicting. For instance, the ability to survive and grow in drying soils is partly
150 determined by physiological properties related to leaf gas exchange, turgor maintenance, and the
151 hydraulic conductivity of water through its roots, stem and leaves (Bartlett et al. 2012). However, such
152 traits are not currently available for most species, necessitating approaches from more widely measured
153 traits for large-scale prediction, such that additional traits can be utilized as they become available.

154 In this study we evaluate four widely measured traits (leaf area, leaf dry matter content, specific
155 leaf area and seed mass) as predictors of Ellenberg indicator scores for light, soil moisture and soil
156 nutrient availability. Average values of each of these traits are already available for large numbers of
157 plant species worldwide, each can be quickly and easily measured with minimal cost, and each has been
158 shown to vary along these environmental gradients for at least some sites and combinations of species
159 as reviewed in Garnier and Navas (2013). For instance, small leaf size has often been associated with
160 dry (Greive 1956, Monk 1966, Scoffoni et al. 2011) and/or nutrient-poor soils (McDonald et al. 2003).
161 Species having large leaves or large SLA are often associated with low light conditions (Hodgson et al.
162 2011). Increasing SLA and decreasing LDMC have been associated with higher soil fertility (Hodgson et

163 al. 2011). Larger seeds have been associated with dry, poorer soils and more shaded habitats
164 (Tautenhahn et al. 2008).

165 **Methods**

166 *Data base for developing prediction equations*

167 The trait data for specific leaf area (SLA, $\text{mm}^2 \text{mg}^{-1}$), leaf dry matter content (LDMC, g g^{-1}), leaf
168 area (LA, mm^2) and seed mass (SM, mg) were previously published values obtained from the TRY
169 database (Kattge et al. 2011), www.try-db.org, accessed 10/07/2013. We calculated the mean value for
170 each species across all observations of the species in TRY. We used Ellenberg indicator scores that can
171 be directly related to resource availability: irradiance (light) level, soil nutrients and moisture, which take
172 integer values from 1 to 9. Although the full range of Ellenberg indicator scores for moisture go to 12,
173 we excluded aquatics, thus limiting this indicator range to 9. Note that Ellenberg (1978) originally
174 described one of his indices an index of soil nitrogen but this index is more properly interpreted as a
175 more general description of soil fertility, soil nutrient availability or site productivity (Hill et al. 1999,
176 Schaffers et al. 2000) and is here called a “nutrient” index. These Ellenberg indicator scores are ordinal
177 scores and their English descriptions are given in Appendix 1. The TRY database contains Ellenberg
178 numbers for 1835 species.

179 We chose four traits that previous authors have associated with one or more of these resources
180 and that were sufficiently documented in the TRY database along with the Ellenberg values: specific leaf
181 area (SLA, $\text{mm}^2 \text{mg}^{-1}$), leaf dry matter content (LDMC, g g^{-1}), leaf area (LA, mm^2) and seed mass (SM, mg).
182 We then calculated the mean value of each trait. Other plant traits from the TRY database that can
183 reasonably be linked to these resource gradients, such as leaf nitrogen concentration, leaf thickness, net
184 photosynthetic rate or leaf lifespan, did not have sufficient coverage with respect to the Ellenberg
185 indicator scores and so were not included in this analysis. For instance, the total number of species with

186 complete data with respect to Ellenberg nutrients goes from 922 using our chosen four traits to 399 if
187 we include leaf nitrogen content and to only 55 if we also include leaf lifespan. Finally, since the
188 relationships between traits and habitat affinities might differ between different basic plant types, we
189 included the “graminoid/herb/shrub/tree” classification in TRY as a categorical factor. A few species,
190 classified as “ferns” were excluded since they do not have a seed size. A few species, ambiguously
191 classified as “herb/shrub” or “shrub/tree”, were classified as “herbs” and “trees” respectively. The
192 number of species having complete coverage for all four traits and each Ellenberg score, as well as the
193 distribution of species among the different Ellenberg scores, are given in Table 1. Of the three Ellenberg
194 scores, the distribution of species for Ellenberg irradiance was the most problematic, since 86% of the
195 981 species had Ellenberg scores of 6 or more.

196 *Statistical analysis and development of the prediction equations*

197 Since the Ellenberg values are only ordered ranks, we used cumulative link models to predict
198 these Ellenberg scores from the four traits. This was done using the `clm` function of the `ordinal` library
199 (Christensen 2015) of R. Cumulative link models are generalizations of maximum likelihood logistic
200 regression to more than two ordered states (Agresti 2002). Given an Ellenberg scale with 9 ordered
201 states ($j=1,9$), the probability that species i will be found in a state less than, or equal to level j , given its
202 trait values is given in Equation 1. Here we present only the simplest version in Equation 1 but this can
203 be generalized to include the “plant form” factor and any combination of interaction terms, just as in
204 any other generalized linear model. LS_i is the latent score for species i , α_j is the intercept, or “threshold
205 value”, i.e. the value of the latent score for which a species has a higher probability of being classified in
206 state j than in a lower state, and β 's are partial slopes. The latent score can be interpreted as the
207 predicted position of a species along the unmeasured latent environmental gradient, measured on a
208 continuous scale as defined by the traits, and which is then divided into ordinal classes as represented

209 by the Ellenberg indicator scores. The probability of a species i being associated with each state j of an
 210 Ellenberg scale is given in Equation 2.

$$211 \quad LS_i = \beta_1 \ln(LA_i) + \beta_2 \ln(LDMC_i) + \beta_3 \ln(SLA_i) + \beta_4 \ln(SM_i) \quad (1a,b)$$

$$p_i(\leq j) = \frac{e^{\alpha_j - LS_i}}{1 + e^{\alpha_j - LS_i}}$$

$$212 \quad p_i(j=1) = p_i(\leq 1)$$

$$p_i(j) = p_i(\leq j) - p(\leq j-1); 1 < j < 9 \quad (2a,b,c)$$

$$p_i(j=9) = 1 - p(\leq 8)$$

213 To use these prediction equations for a new species, once parameterized, one would first enter
 214 its trait values in equation 1a to get the latent score (LS_i), then enter this latent score into equation 1b to
 215 obtain the probability that this species has an Ellenberg indicator value of j or less. To obtain the
 216 probability that the species has an Ellenberg indicator value of exactly j , one would apply equations
 217 2a,b,c. A worked example is provided in Appendix 1. These calculations are automated using the
 218 “predict” function associated with the `clm` function in R. Although a very large number of possible
 219 models can be generated given four traits plus the categorical “plant type” including all possible
 220 interactions, we present two for each Ellenberg indicator variable: (i) a simplified model involving only
 221 the four traits as main effects for each of the four plant types and (ii) the best model, based on AIC
 222 values and obtained using the “stepAIC” function in the MASS library of R (Venables et al. 1994). Given
 223 the large sample sizes and resulting statistical power, this second model was invariably more
 224 complicated and involved several higher-order interaction terms.

225 *Evaluating predictive error*

226 The above equations only predict the probability that any species will be classified into each
 227 level of a given Ellenberg scale. By multiplying each level ($j=1$ to 9) to the predicted probability (p_{ij}) of a

228 species i being classified into each level, i.e. ($\hat{s}_i = \sum_j j \cdot p_{ij}$) we get the mean predicted Ellenberg score
 229 for each species, which varies continuously from 1 to 9. To determine how well the average predicted
 230 score of the species agrees with the actual scores (s_i), we estimated the distribution of errors as the
 231 percentage of species having a given difference between the observed (s_i) and predicted scores and the
 232 mean predictive error (MPE, Equation 3) of cross-validated models (see below), which measures the
 233 average error (number of Ellenberg levels) between the observed (s) and predicted scores.

$$234 \quad MPE = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{s}_i)^2}{n}} \quad (1a,b)$$

235 *Independent validation of the prediction equations*

236 We evaluated the predictive ability of our equations with respect to independent observations
 237 in two ways. First, we used cross-validation; i.e. we randomly chose 80% of our data to fit the equation
 238 and then used these parameterized equations to predict the mean Ellenberg scores of the remaining
 239 20%, after which we calculated the mean predictive error. We did this 100 times, each time randomly
 240 subdividing the data into the calibration and evaluation subsets. This determines the likely error rate
 241 when extending the equations to new species whose Ellenberg values are known.

242 However, the main objective of the study was to extend our predictive ability to species that
 243 were never classified by Ellenberg, including species that are found outside of central Europe. We
 244 therefore identified species for which we had trait data but for which Ellenberg values did not exist, and
 245 for which we could obtain habitat descriptions. Such habitat descriptions are very approximate
 246 (“understory species”, “typically found in wetlands” etc.) and not as detailed as those of Ellenberg. We
 247 therefore concentrated on species with clearly different habitat affinities. First, we identified 1134
 248 species in the TRY database for which there was information on all four traits but for which Ellenberg
 249 scores were missing. Of these 1134 new species, we then identified 55 species, mostly in North

250 America, for which available habitat descriptions (local floras, field guides, online sites such as that
251 managed by the United States Department of Agriculture) clearly indicate that they are most commonly
252 found in wet, but not permanently inundated, soils (approximately Ellenberg moisture values of >6) and
253 43 species that are most commonly found in shaded or partially shaded habitats (approximately
254 Ellenberg irradiance values of <5). Next, using the same 1134 species from the TRY database, we
255 combined this with the BioFlor data base (Klotz et al. 2002; www.ufz.de/bioflor). We used the BioFlor
256 information on “phytosociological affinity” of the plant species following Schubert et al. (2001) and
257 classified each of the phytosociological classes (i.e. habitat types) either as “wet” or “shaded” with
258 values “yes”, “no”, “indifferent” or “unknown”, respectively. Then we simply counted the number of
259 “yes” and “no” occurrences of the species in the habitat types, because species can occur in several
260 habitats. Knowing the species from extensive field experience, classifying species with more “yes” than
261 “no” counts for “wet” habitats and more or equal counts of “yes” compared to “no” for shaded habitats
262 proved to yield sensible results in classifying them. This gave us a further 55 and 296 species classified as
263 occurring or not occurring on wet soils, 87 and 264 species classified as occurring or not occurring in
264 shaded habitats. Finally, we used the data set of Frenette-Dussault et al. (2013) consisting of 34 species
265 of the Moroccan steppe, close to the Saharan desert, which has very dry and nutrient-poor soils and
266 uniformly low and sparse vegetation growing in full sun. The predicted mean Ellenberg scores (Eqn. 1)
267 for all of these species with respect to soil moisture and light were then calculated.

268 If the prediction equations can be generalised then the predicted mean Ellenberg scores of
269 these species with respect to Ellenberg moisture should be lowest for the Moroccan species, be higher
270 for the TRY “non-wetland” species, and be highest for the “wetland” species. If the prediction
271 equations can be generalised then the mean Ellenberg scores with respect to irradiance should be
272 lowest for the “shade” species, be larger for the “non-shade” species, and be highest for the Moroccan
273 species. Finally, the mean Ellenberg nutrient scores should be lower in the Moroccan steppe species

274 than in the other species. Since the mean predicted Ellenberg scores for both moisture and light were
275 approximately normally distributed, we tested the hypothesised differences in the latent scores using
276 one-way ANOVA followed by Tukey post-hoc tests.

277 **Results**

278 Table 2 presents the cumulative link models. Two models are presented for each gradient: a
279 “simplified” model (i.e. each trait without interactions between them, but including differences in slopes
280 between plant types) and the model having the best AIC score, which is always more complicated. All
281 coefficients in these models, except for the partial slope associated with seed mass for the Ellenberg
282 nutrient scores, are significantly different from zero ($p < 0.05$). The average predictive error of these
283 models, based on 100 independent cross-validation runs, varied between 1.3 (light) to 1.8 (nutrients,
284 moisture) Ellenberg ranks. Table 3 presents the distribution of errors for each model and Figure 1 shows
285 the predicted vs. observed cross-validated values for the AIC-selected models. Between 70% (soil
286 fertility and moisture) and 90% (irradiance) of the cross-validated predictions were within one Ellenberg
287 rank of the actual value and at least 90% were within two Ellenberg ranks. The prediction errors for the
288 simplified models were very similar to the more complicated models, showing that the more
289 complicated models, while capturing statistically significant effects, provided only minor improvements
290 in predictive ability.

291 In order to illustrate the trait – environment relationships predicted by the models, we plotted
292 the predicted Ellenberg scores as a function of the levels of variation in each trait for each plant type
293 while fixing the other traits at their mean values. The results are shown in Figures 2 – 4. In general,
294 both increasing leaf size and decreasing LDMC (i.e. less dense leaf tissues) indicated habitat preferences
295 for soils that were wetter and with more nutrients, i.e. more productive habitats. These two traits did
296 not respond strongly to changing irradiance levels. The only exceptions were in the trees, whose LDMC

297 values increased with increasing habitat productivity and the graminoids, whose leaf areas decreased
298 with increasing light levels.

299 Increasing SLA (panel C of figures 2-4) coincide with an affinity for habitats having wetter, more
300 fertile soils and also lower light levels. Two partial exceptions were seen in the graminoid species and
301 the trees. SLA in the graminoid species and trees was relatively unresponsive to soil nutrient levels and
302 SLA actually decreased with increasing moisture levels in the graminoids.

303 Seed size was the least informative trait with respect to the three Ellenberg gradients. It
304 decreased with increasing soil moisture (except for the herbs) but was largely unresponsive to soil
305 nutrients and decreased only slightly (and not at all for the herbs and shrubs) with increasing light levels.

306 Next, we predicted the mean Ellenberg scores for those species that were never classified by
307 Ellenberg. Of the 1134 species in the TRY data base for which Ellenberg numbers were missing but for
308 which we had values for the four traits, we were able to classify 125 species as commonly occurring in
309 “wet” soils and 264 species as commonly occurring in drier soils (i.e. not “wet” habitats). Similarly, we
310 could classify 91 species as commonly occurring in “shaded” habitats and 296 species as commonly
311 occurring in non-shaded habitats (i.e. “open” habitats). We further included the 34 species that occur in
312 the Moroccan steppe, which is a very dry, nutrient poor and open habitat, giving a total of 423 and 421
313 species classified with respect to soil wetness and habitat light levels, respectively. The average
314 predicted Ellenberg scores for soil moisture increased as predicted (Figure 5): Moroccan species < dry
315 species (TRY) < wet species (TRY). These mean Ellenberg scores differed significantly ($F_{2,370}=15.40$,
316 $p=3.7e^{-07}$) and each was different from the others based on a Tukey post-hoc test. The average predicted
317 Ellenberg scores for light also increased as predicted (Figure 5): Moroccan species > open species (TRY) >
318 shade (forest) species (TRY). These mean Ellenberg scores differed significantly ($F_{2,370}=31.99$, $p=2.2e^{-16}$)
319 and each was different from the others based on a Tukey post-hoc test. Finally, the average predicted

320 Ellenberg scores for nutrients were significantly lower in the Moroccan steppe species than for those of
321 the other species ($F_{1,371}=80.62$, $p= 2.2e^{-16}$) .

322 **Discussion**

323 Trait – environment relationships

324 Our models predicted that herbaceous species having very small, dense leaves with a low SLA
325 plus (to a lesser degree) larger seeds are more likely be found on the driest and the most infertile soils.
326 In fact, the models for soil moisture and soil nutrients (Ellenberg's N) were quite similar even though the
327 two sets of Ellenberg scores are only moderately correlated (Spearman $r=0.36$). Although these
328 qualitative trait – environment trends have been repeatedly reported in the literature, the trees – and
329 especially the graminoids - responded rather differently. Increasing SLA in these species was indicative
330 of dryer sites (decreased Ellenberg moisture) but changed little with respect to habitat soil nutrients
331 (Ellenberg N). A possible explanation can be found by decomposing SLA into its components (Vile et al.
332 2005). SLA is approximately equal to $1/(LDMC*T)$ (Shiple et al. 2002) for typical laminar leaves, where
333 T is leaf lamina thickness; it follows that $\ln(SLA)\sim-\ln(LDMC)-\ln(T)$. The negative relationship between
334 SLA and LDMC can therefore be modulated by changes in lamina thickness and it has already been
335 shown that, although graminoids and herbs have largely overlapping ranges of SLA, the relative
336 importance of LDMC vs. T in determining this range is quite different (Pyankov et al. 1999). With respect
337 to increasing Ellenberg nutrients, graminoids decreased LDMC (which would increase SLA) but must
338 have also increased lamina thickness (which would decrease SLA). As a result, SLA was largely
339 insensitive to habitat nutrient levels. With respect to increasing Ellenberg moisture, graminoids did not
340 change LDMC and so the increased lamina thickness actually decreased SLA.

341 The main response to increasing light levels in the herbs was to decrease SLA even though leaf
342 size and LDMC responded only slightly. Again, a large decrease in SLA without a correspondingly large

343 increase in LDMC means that the decrease in SLA in the herbs was mostly driven by increasing lamina
344 thickness as irradiance levels increased. For the graminoids and trees, the decrease in SLA with
345 increasing light levels was accompanied by an increase in LDMC, and so lamina thickness for these plant
346 types did not change as much. The response of the trees (and somewhat for the shrubs) to levels of
347 habitat irradiance should be interpreted with care since most measurements of SLA in trees are taken
348 on sun leaves. These differences in response between plant types are further complicated by the fact
349 that the number herb species (691) was much greater than that of the shrubs (44), trees (31) or
350 graminoids (156) and the range of trait variation in the plant types other than the herbs was often quite
351 limited, as seen in Figures 1 -3.

352 Predictive accuracy and generality

353 Through cross-validation we predicted Ellenberg scores that were within 2 ranks of the observed
354 values in over 90% of the cases. Given that the ranks range from 1 to 9, this means that we could
355 distinguish between low, medium and high scores but not within adjacent ranks. In 40-64% of the cases
356 (depending on the Ellenberg characteristic), the exact score was predicted, with Ellenberg irradiance
357 having the highest accuracy. However, these summary statistics do not provide a complete view of the
358 predictive accuracy of our models.

359 First, the increased predictive accuracy for Ellenberg irradiance is a consequence of the fact that
360 most species had Ellenberg irradiance scores of 6 or more (plants generally occurring in well-lit places or
361 in only partial shade). As Figure 1 shows, our equations do a poor job of differentiating between
362 Ellenberg irradiance levels of 5 or less. In part this is because so few species had such scores (Table 1)
363 that such species had little weight in the model estimation. However, another reason is because there
364 might actually be little real difference in these first five ranks, at least with respect to the chosen traits.
365 An Ellenberg irradiance rank of 5 is a “semi-shade plant, rarely in full light, but generally with more than

366 10% relative illumination when trees are in full leaf". In other words, ranks 1 -5 all indicate species that
367 are generally found in forested understories.

368 Second, as Figure 1 shows, our equations predict values that are higher than observed for the
369 lowest Ellenberg scores and values that are lower than observed for the highest Ellenberg scores. In
370 part, this is due to low representation of species at these extremes (Table 1) but it is also due to an
371 unavoidable mathematical bias. Since these are ordinal scales that are bounded by 1 and 9, any
372 prediction error at these bounds must increase the predicted values at the lower bound and decrease
373 them at the upper bound and this will necessarily generate such a bias. One should therefore use Figure
374 1 to modify predicted Ellenberg values.

375 There are some obvious sources of error in these data. First, a substantial proportion of the
376 unexplained variation is probably due the arbitrary nature of the original Ellenberg values, leading to
377 errors in classification of the habitat affinities of the species. Second, the Ellenberg indicator scores do
378 not include information on niche breadth. Although Ellenberg's classification was based on a lifetime of
379 experience, the habitat descriptions for each class are sufficiently vague, and the field distributions of
380 many species with respect to these habitat descriptions are sufficiently difficult to determine, that many
381 species could probably be assigned to several adjacent Ellenberg classes with equal confidence. It is
382 therefore not surprising that our prediction equations cannot discriminate between adjacent Ellenberg
383 classes. Third, the trait values used in the analysis are estimates of species' means obtained from a
384 heterogeneous collection of values uploaded to the TRY database; we do not know how closely the
385 species' trait means calculated from the TRY database correspond to the true species' trait means
386 (Cordlandwehr et al. 2013) and, of course, intraspecific trait variation is completely ignored. A necessary
387 consequence of errors in the predictor variables (i.e. the species' means for each trait) is also to reduce
388 the predictive ability of the model. Finally, these equations treat each environmental gradient

389 separately but these are correlated in nature and the actual habitat affinities of species would reflect
390 correlated trait adaptations to all environmental gradients. Beyond these sources of error, it is surely
391 true that more (or better) traits would further improve the predictive ability of our equations

392 The main goal of this research was to determine the degree to which our easily measured and
393 widely available trait values could be used to predict the habitat affinities of species not already
394 classified by Ellenberg values and, especially, species for which no habitat affinities are available. Can
395 one use these equations to predict the Ellenberg habitat affinities of species not classified by Ellenberg,
396 especially of species outside of Central Europe? Strictly speaking, it is impossible answer this question
397 since these numbers represent Ellenberg's expert opinion rather than some independently measurable
398 attribute. However, if our equations are generalizable, we should be able to correctly differentiate
399 between species that typically occupy clearly different points along the underlying environmental
400 gradients. Our results (Figure 5) suggest that this is possible. With respect to soil fertility our equations
401 assigned the Moroccan steppe species to Ellenberg levels of 1 ("extremely infertile sites") to 3 ("more or
402 less infertile sites") while the heterogeneous collection of other species had a median score of around 5
403 ("sites with intermediate fertility"). With respect to soil moisture and irradiance our equations
404 performed more poorly. Although the order of the predicted ranks for the "wetland", "non-wetland"
405 and "steppe" species was correct, the median moisture score for the "wetland" species was ~5.5 when it
406 should have been around 7 ("plants typically found on constantly moist or damp") or more. This level of
407 error was also seen in the cross-validated estimates (Figure 1). Similarly, although the order of the
408 predicted ranks for the "steppe", "open" and "forest" species was in correct, and while the "steppe" and
409 "open habitat" species were assigned reasonable irradiance ranks, the species described as understory
410 plants were incorrectly assigned Ellenberg ranks of between 6 and 7 when they should have received
411 ranks of 5 or less. Our equations did particularly poorly when identifying understory species (Figure 1).

412 We therefore propose our equations as a general, but only very approximate, method of
413 describing the likely habitat affinities of soil moisture, soil nutrients and irradiance levels for species
414 lacking such information. When distribution models already exist that include information on the
415 relevant abiotic gradients for a given species then these would likely provide better predictions than
416 would our equations. Species distribution models (Elith et al. 2009, Peterson 2011) provide an alternate
417 approach for estimating the realized niche of species. Because of the growing availability of species
418 occurrence data and environmental layers (Maldonado et al. 2015), habitat models can now be readily
419 constructed for many species. However, these models tend to be more appropriate at larger spatial
420 scales and often describe only large-scale climate variables like temperature and precipitation.
421 Currently, such distribution models exist for only a tiny fraction of plant species. Since it is much easier
422 to obtain information on a few static traits than to construct a species distribution model, we expect
423 that our prediction equations will still be useful for the majority of species. Ultimately, as trait
424 databases like TRY continue to accumulate trait information on more species, one may be able to use
425 our equations to product coarse habitat descriptions for many thousands of species.

426 **Acknowledgements**

427 This research was partially funded by a Natural Sciences and Engineering Research Grant to BS. VO was
428 supported by the Russian Science Foundation (# 14-50-00029). The study has been supported by the
429 TRY initiative on plant traits (<http://www.try-db.org>). The TRY initiative and database is hosted, at the
430 Max Planck Institute for Biogeochemistry, Jena, Germany. TRY is currently supported by
431 DIVERSITAS/Future Earth and the German Centre for Integrative Biodiversity Research (iDiv) Halle-
432 Jena-Leipzig.

433 **References**

- 434 Agresti, A. 2002. *Categorical Data Analysis*. John Wiley & Sons, NY.
- 435 Bartelheimer, M. & Poschlo, P. 2014. The response of grassland species to nitrate versus ammonium
436 coincides with their pH optima. *Journal of Vegetation Science* 25: 760-770.
- 437 Bartlett, M.K., Scoffoni, C., Ardy, R., Zhang, Y., Sun, S., Cao, K. & Sack, L. 2012. Rapid determination of
438 comparative drought tolerance traits: Using an osmometer to predict turgor loss point. *Methods in
439 Ecology and Evolution* 3: 880-888.
- 440 Carpenter, W. & Goodenough, A. 2014. How robust are community-based plant bioindicators? Empirical
441 testing of the relationship between Ellenberg values and direct environmental measures in woodland
442 communities. *Community Ecology* 15: 1-11.
- 443 Christensen, R.H.B. 2015. ordinal - Regression Models for Ordinal Data. R package version 2015.6-28. In:
444 <http://www.cran.r-project.org/package=ordinal/>.
- 445 Cordlandwehr, V., Meredith, R.L., Ozinga, W.A., Bekker, R.M., Van Groenendael, J.M. & Bakker, J.P.
446 2013. Do plant traits retrieved from a database accurately predict on-site measurements? *Journal of
447 Ecology* 101: 662-670.
- 448 Douma, J.C., Witte, J.-P.M., Aerts, R., Bartholomeus, R.P., Ordoñez, J.C., Venterink, H.O., Wassen, M.J. &
449 van Bodegom, P.M. 2012. Towards a functional basis for predicting vegetation patterns; incorporating
450 plant traits in habitat distribution models. *Ecography* 35: 294-305.
- 451 Elith, J. & Leathwick, J.R. 2009. Species distribution models: Ecological explanation and prediction across
452 space and time. In: *Annual Review of Ecology, Evolution, and Systematics*, pp. 677-697.
- 453 Ellenberg, H. 1988. *Vegetation Ecology of Central Europe*. 4th. Cambridge University Press, Cambridge,
454 UK.
- 455 Ellenberg, H. 1978. *Vegetation Mitteleuropas mit den Alpen in ökologischer Sicht*. Verlag.
- 456 Ellenberg, H. 1992. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18: 1-258.
- 457 Ertsen, A.C.D., Alkemade, J.R.M. & Wassen, M.J. 1998. Calibrating Ellenberg indicator values for
458 moisture, acidity, nutrient availability and salinity in the Netherlands. *Plant Ecol.* 135: 113-124.
- 459 Franzaring, J., Fangmeier, A. & Hunt, R. 2007. On the consistencies between CSR plant strategies and
460 Ellenberg ecological indicator values. *Journal of Applied Botany and Food Quality* 81: 86-94.
- 461 Frenette-Dussault, C., Shiple, B., Meziane, D. & Hingrat, Y. 2013. Trait-based climate change predictions
462 of plant community structure in arid steppes. *Journal of Ecology* 101: 484-492.
- 463 Fujita, Y., Van Bodegom, P.M. & Witte, J.M. 2013. Relationships between nutrient-related plant traits
464 and combinations of soil N and P fertility measures. *PLoS ONE* 8.
- 465 Garnier, E. & Navas, M.L. 2013. *Diversité fonctionnelle des plantes. Traits des organismes, structure des
466 communautés, propriétés des écosystèmes*. de boeck.
- 467 Greive, B.J. 1956. Studies in the water relations of plants: I. Transpiration of western Australian (Swan
468 Plain) sclerophylls. *Royal Society of Western Australia* 40: 15-20.
- 469 Grime, J.P. 1979. *Plant strategies and vegetation processes*. John Wiley and Sons, New York.
- 470 Guarino, R., Domina, G. & Pignatti, S. 2012. Ellenberg's Indicator values for the Flora of Italy – first
471 update: Pteridophyta, Gymnospermae and Monocotyledoneae. *Flora Mediterranea* 22: 197-209.
- 472 Hill, M.O., Mountford, J.O., Roy, D.B. & Bunce, R.G.H. 1999. Ellenberg's Indicator Values for British
473 Plants. In: Institute of Terrestrial Ecology, Huntingdon, UK.
- 474 Hodgson, J.G., Montserrat-Martí, G., Charles, M., Jones, G., Wilson, P., Shiple, B., Sharafi, M.,
475 Cerabolini, B.E.L., Cornelissen, J.H.C., Band, S.R., Bogard, A., Castro-Díez, P., Guerrero-Campo, J., Palmer,
476 C., Pérez-Rontomé, M.C., Carter, G., Hynd, A., Romo-Díez, A., De Torres Espuny, L. & Royo Pla, F. 2011. Is
477 leaf dry matter content a better predictor of soil fertility than specific leaf area? *Annals of Botany* 108:
478 1337-1345.

- 479 Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:
480 415–427.
- 481 Karrer, G. 1992. Österreichische Waldboden-Zustandsinventur. Teil VII: Vegetationsökologische
482 Analysen- Mitt. Forstl. *Bundesversuchsanst. Wien* 168: 193-242.
- 483 Kattge, J., Diaz, S., Lavorel, S., Prentice, C., Leadley, P., Bonisch, G., Garnier, E., Westoby, M., Reich, P.B.,
484 Wright, I.J., Cornelissen, J.H.C., Violle, C., Harrison, S.P., van Bodegom, P.M., Reichstein, M., Enquist, B.J.,
485 Soudzilovskaia, N.A., Ackerly, D.D., Anand, M., Atkin, O., Bahn, M., Baker, T.R., Baldocchi, D., Bekker, R.,
486 Blanco, C.C., Blonder, B., Bond, W.J., Bradstock, R., Bunker, D.E., Casanoves, F., Cavender-Bares, J.,
487 Chambers, J.Q., Chapin, F.S., Chave, J., Coomes, D., Cornwell, W.K., Craine, J.M., Dobrin, B.H., Duarte, L.,
488 Durka, W., Elser, J., Esser, G., Estiarte, M., Fagan, W.F., Fang, J., Fernandez-Mendez, F., Fidelis, A.,
489 Finegan, B., Flores, O., Ford, H., Frank, D., Freschet, G.T., Fyllas, N.M., Gallagher, R.V., Green, W.A.,
490 Gutierrez, A.G., Hickler, T., Higgins, S.I., Hodgson, J.G., Jalili, A., Jansen, S., Joly, C.A., Kerkhoff, A.J.,
491 Kirkup, D., Kitajima, K., Kleyer, M., Klotz, S., Knops, J.M.H., Kramer, K., Kuhn, I., Kurokawa, H., Laughlin,
492 D., Lee, T.D., Leishman, M., Lens, F., Lenz, T., Lewis, S.L., Lloyd, J., Llusia, J., Louault, F., Ma, S., Mahecha,
493 M.D., Manning, P., Massad, T., Medlyn, B.E., Messier, J., Moles, A.T., Muller, S.C., Nadrowski, K., Naeem,
494 S., Niinemets, U., Nollert, S., Nuske, A., Ogaya, R., Oleksyn, J., Onipchenko, V.G., Onoda, Y., Ordonez, J.,
495 Overbeck, G., Ozinga, W.A., Patino, S., Paula, S., Pausas, J.G., Penuelas, J., Phillips, O.L., Pillar, V., Poorter,
496 H., Poorter, L., Poschlod, P., Prinzing, A., Proulx, R., Rammig, A., Reinsch, S., Reu, B., Sack, L., Salgado-
497 Negre, B., Sardans, J., Shiodera, S., Shipley, B., Siefert, A., Sosinski, E., Soussana, J.F., Swaine, E.,
498 Swenson, N., Thompson, K., Thornton, P., Waldram, M., Weiher, E., White, M., White, S., Wright, S.J.,
499 Yguel, B., Zaehle, S., Zanne, A.E. & Wirth, C. 2011. TRY - a global database of plant traits. *Global Change*
500 *Biology* 17: 2905-2935.
- 501 Klaus, V.H., Kleinebecker, T., Boch, S., Müller, J., Socher, S.A., Prati, D., Fischer, M. & Hölzel, N. 2012.
502 NIRS meets Ellenberg's indicator values: Prediction of moisture and nitrogen values of agricultural
503 grassland vegetation by means of near-infrared spectral characteristics. *Ecological Indicators* 14: 82-86.
- 504 Klotz, S., Kühn, I. & Durka, W. 2002. Biolflor - eine Datenbank zu biologisch-ökologischen Merkmalen zur
505 Flora von Deutschland. *Schriftenreihe für Vegetationskunde* 28: 1-333.
- 506 Landolt, E. 1977. Ökologische Zeigerwerte zur Schweizer Flora. *Veröff. Geobot. Inst. Eidg. Techn.*
507 *Hochsch. Stift. Rübel Zür.* 68: 1-208.
- 508 Maldonado, C., Molina, C.I., Zizka, A., Persson, C., Taylor, C.M., Albán, J., Chilquillo, E., Rønsted, N. &
509 Antonelli, A. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent
510 can we trust public databases? *Global Ecology and Biogeography* 24: 973-984.
- 511 McDonald, P.G., Fonseca, C.R., Overton, J.M. & Westoby, M. 2003. Leaf-size divergence along rainfall
512 and soil-nutrient gradients: Is the method of size reduction common among clades? *Functional Ecology*
513 17: 50-57.
- 514 Monk, C.D. 1966. An ecological significance of evergreenness. *Ecology* 47: 504-505.
- 515 Ozinga, W.A., Schaminée, J.H.J., Bekker, R.M., Bonn, S., Poschlod, P., Tackenberg, O., Bakker, J. &
516 Groenendael, J.M.v. 2005. Predictability of plant species composition from environmental conditions is
517 constrained by dispersal limitation. *Oikos* 108: 555-561.
- 518 Peterson, A.T. 2011. *Ecological niches and geographic distributions*. Princeton University Press,
519 Princeton.
- 520 Pignatti, S. 2005. Valori di bioindicazione delle piante vascolari della Flora d'Italia. *Braun-Blanquetia* 39:
521 3-95.
- 522 Pyankov, V., Kondratchuk, A. & Shipley, B. 1999. Leaf structure and specific leaf mass: the alpine desert
523 plants of the Eastern Pamirs (Tadjikistan). *The New Phytologist* 143: 131-142.
- 524 Schaffers, A.P. & Sykora, K.V. 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and
525 soil reaction: a comparison with field measurements. *Journal of Vegetation Science* 11: 225-244.

- 526 Schubert, R., Hilbig, W. & Klotz, S. 2001. *Bestimmungsbuch der Pflanzengesellschaften Deutschlands*.
527 Spektrum, Heidelberg.
- 528 Scoffoni, C., Rawls, M., McKown, A., Cochard, H. & Sack, L. 2011. Decline of leaf hydraulic conductance
529 with dehydration: relationship to leaf size and venation architecture. *Plant Physiol* 156: 832-843.
- 530 Shipley, B. & Vu, T.-T. 2002. Dry matter content as a measure of dry matter concentration in plants and
531 their parts. *New Phytologist* 153: 359-364.
- 532 Tautenhahn, S., Heilmeier, H., Götzenberger, L., Klotz, S., Wirth, C. & Kühn, I. 2008. On the biogeography
533 of seed mass in Germany – distribution patterns and environmental correlates. *Ecography* 31: 457-468.
- 534 Valladares, F. & Niinemets, Ü. 2008. Shade tolerance, a key plant feature of complex nature and
535 consequences. In: *Annual Review of Ecology, Evolution, and Systematics*, pp. 237-257.
- 536 Venables, W.N. & Ripley, B.D. 1994. *Modern applied statistics with S-PLUS*. Springer-Verlag, New York.
- 537 Vile, D., Garnier, E., Shipley, B., Laurent, G., Navas, M.L., Roumet, C., Lavorel, S., Diaz, S., Hodgson, J.G.,
538 Lloret, F., Midgley, G.F., Poorter, H., Rutherford, M.C., Wilson, P.J. & Wright, I.J. 2005. Specific leaf area
539 and dry matter content estimate thickness in laminar leaves. *Ann Bot* 96: 1129-1136.
- 540 Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E. 2007. Let the concept
541 of trait be functional! *Oikos* 116: 882-892.
- 542 Warming, E. & Vahl, M. 1909. *Oecology of plants - an introduction to the study of plant communities*.
543 Clarendon Press, Oxford.
- 544 Zarzycki, K., Trzcńska-Tacik, H., Różański, W., Szeląg, Z., Wołek, J. & Korzeniak, U. 2002. *Ecological*
545 *indicator values of vascular plants of Poland/Ekologiczne Liczby Wskaźnikowe Roślin Naczyniowych*
546 *Polski*.
- 547
- 548 Appendix 1. Ellenberg scores for irradiance level ("light"), moisture and soil fertility and a worked
549 example of calculations.
- 550

551 Table 1. The number of species (% of total) having complete information on the four traits and for
 552 which the Ellenberg scores for nutrients, moisture and light are known.

553

Ellenberg score	Soil nutrients	Habitat moisture	Irradiance level
1	48 (5.2%)	2 (0.2%)	1 (0.1%)
2	141 (15.3%)	36 (3.6%)	6 (0.6%)
3	127 (13.8%)	130 (13.2%)	19 (1.9%)
4	128 (13.9%)	214 (21.7%)	60 (6.1%)
5	141 (15.3%)	228 (23.1%)	51 (5.2%)
6	111 (12.0%)	107 (10.8%)	141 (14.4%)
7	118 (12.8%)	88 (8.9%)	317 (32.3%)
8	85 (9.2%)	92 (9.3%)	276 (28.1%)
9	23 (2.5%)	91 (9.2%)	110 (11.2%)
Total	922	988	981

554

555 Table 2: Results of the cumulative link model (Equations 1 and 2), giving the probabilities that a species
 556 with given ln-transformed values of leaf area (LA, mm²), leaf dry matter content (LDMC, g g⁻¹), specific
 557 leaf area (SLA, mm² mg⁻¹) and seed mass (SM, mg) would be classified in each of the nine ordinal
 558 Ellenberg classes for site irradiance level (light), site soil nutrients, and soil moisture level. Shown are
 559 the maximum likelihood values for each of four plant types (H= herb, G=graminoid, S=shrub, T=tree. of
 560 the partial slopes associated with each of the four ln-transformed traits and their interactions which
 561 together give the latent score. Also shown are the “threshold values” or intercepts; thus “x|y” (eg. 1|2)
 562 gives the value of the latent score, calculated from the partial slopes and trait values, at which a species
 563 is more likely to be in Ellenberg state 2 rather than 1. Two sets of models are shown: a simplified model
 564 that doesn’t include interactions between traits and the best model as determined by AIC values. MPE
 565 gives the mean predictive error for each model.

Simplified model without interactions												
Term	Ellenberg nutrients				Ellenberg light				Ellenberg soil moisture			
	G	H	S	T	G	H	S	T	G	H	S	T
intercept	0	0.5	0.47	3.83	0	-3.57	-8.36	-8.41	0	7.14	3.95	3.23
ln(LA)	0.67	0.44	0.59	0.38	-0.77	-0.17	-0.21	-0.75	0.16	0.21	0.06	0.37
ln(LDMC)	1.52	0.65	0.04	1.33	-0.48	-0.39	-2.84	1.03	0.48	0.18	0.41	0.53
ln(SLA)	0.73	1.4	1.45	0.82	-2.16	-2.15	-1.41	0.9	0.63	1.32	1.22	0.42
ln(SM)	0.1	0.09	0.14	0.12	0.04	-0.09	-0.18	-0.08	-0.3	0.15	0.14	0.17
Threshold coefficients												
1 2	5.27				-18.41				-7.41			
2 3	7.2				-16.45				-4.41			
3 4	8.12				-15.09				-2.68			
4 5	8.88				-13.72				-1.43			
5 6	9.68				-13.11				-0.37			
6 7	10.37				-12.05				0.18			
7 8	11.41				-10.41				0.74			
8 9	13.18				-8.45				1.6			
AIC	3568.02				3061.75				3773.2			
MEP	1.81				1.3				1.8			
Best model using AIC												

Term	Ellenberg nutrients				Ellenberg light				Ellenberg soil moisture			
	G	H	S	T	G	H	S	T	G	H	S	T
intercept	0	2.13	6.95	1.13	0	59.57	42.52	173.73	0	9.15	7.05	4.32
A=ln(LA)	0.82	0.04	0.66	1.06	8.73	0.17	3.46	17.9	1.84	1.84	1.84	1.84
B=ln(LDMC)	8.94	5.12	0.48	9.22	42.12	3.27	4.05	127.52	3.82	3.99	4.66	4.18
C=ln(SLA)	0.25	1.66	1.96	0.54	19.52	-2.04	3.74	-33.31	4.94	2.35	2.31	3.18
D=ln(SM)	0.52	0.52	0.52	0.52	-28.3	3.54	11.51	-41.67	0.08	0.38	0.25	0.05
A*B	0.64	0.17	0.46	1.04	6.12	-0.51	-0.03	-17.34	0.94	0.94	0.94	0.94
A*C	0.24	0.24	0.24	0.24	-3.36	-0.11	-1.44	4.94	0.7	0.7	0.7	0.7
A*D	0.04	0.04	0.04	0.04	4.02	-0.98	0.47	5.57	0.07	0.07	0.07	0.07
B*C	1	1	1	1	14.5	-0.8	-1.55	-38.38	1.56	1.56	1.56	1.56
B*D	0.24	0.24	0.24	0.24	26.05	0.95	18.64	-35.12	1.32	1.63	0.95	1.02
C*D	0.19	0.19	0.19	0.19	8.39	-1.26	3.65	13.17	0.42	0.42	0.42	0.42
A*B*C					-2.16	0.4	-0.1	5.3	0.32	0.32	0.32	0.32
A*B*D					3.52	-0.43	1.7	4.48	0.08	0.08	0.08	0.08
A*C*D					-1.14	0.33	-0.15	-1.79	0.08	0.08	0.08	0.08
B*C*D					7.74	-0.33	6.11	11.39	0.27	0.27	0.27	0.27
A*B*C*D					-1	0.15	-0.56	-1.46				
Threshold coefficients												
1 2	9.54				43				-19.7			
2 3	11.5				44.96				-16.59			
3 4	12.43				46.34				-14.8			
4 5	13.2				47.83				-13.52			
5 6	14.01				48.5				-12.43			
6 7	14.72				49.62				-11.86			
7 8	15.76				51.36				-11.29			
8 9	17.55				53.42				-10.4			
AIC	3559.16				3054.74				3744.86			
MEP	1.77				1.27				1.76			

566

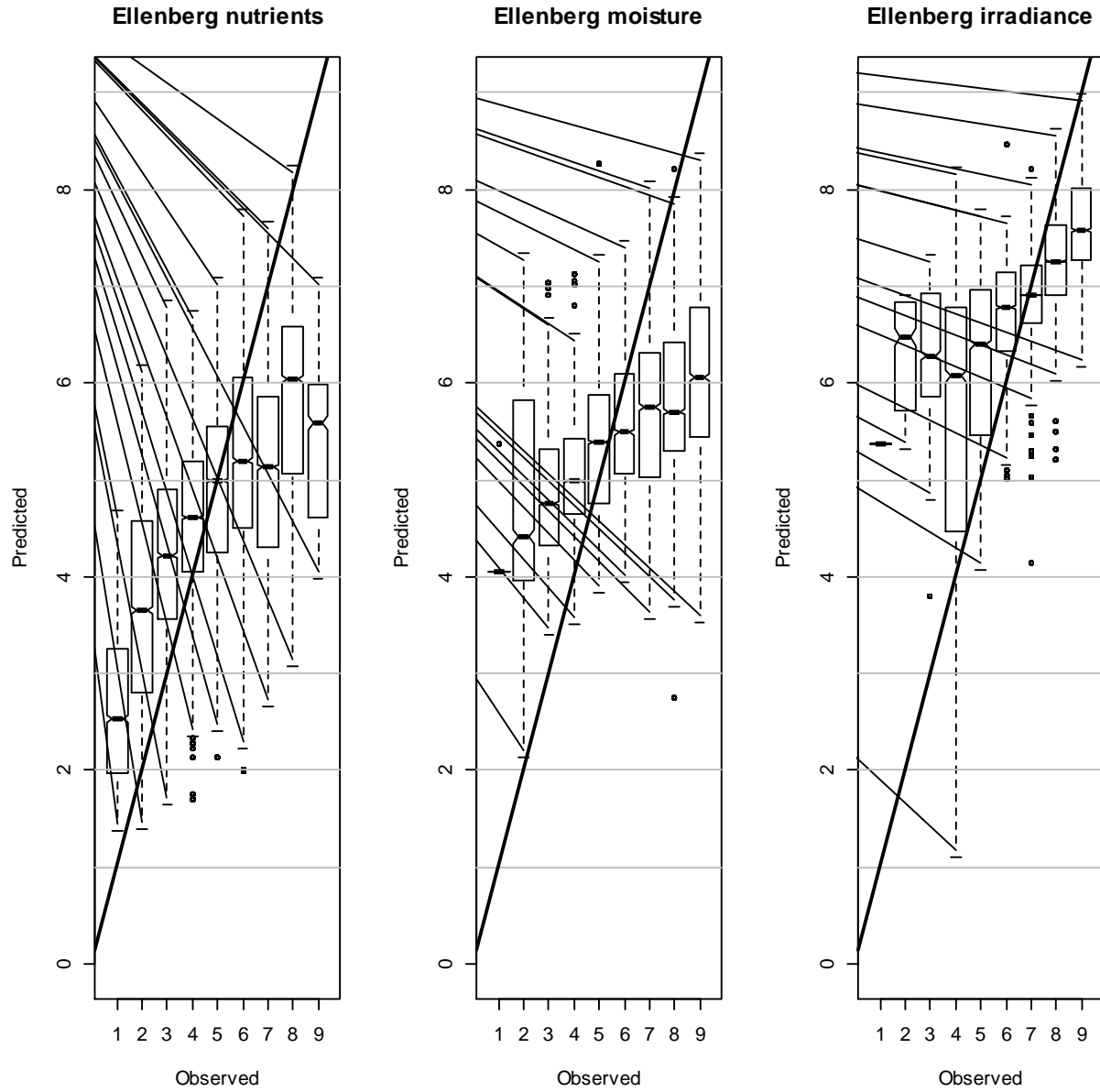
567

568 Table 3. Levels of prediction error in cross-validated data (100 independent runs, 20% of data for each
 569 run) for the Ellenberg scores of soil nutrients, habitat moisture, and irradiance, each having 9 levels. The
 570 first column (“Error” is the difference in the observed and predicted Ellenberg score; thus, 1=the
 571 predicted score is within 1 score of the real score, etc. Subsequent columns give the percent of the
 572 cross-validated observations having this level of prediction error (cumulative percent error). These error
 573 rates are given for both the AIC-selected models and the simplified models.

Error	soil nutrients		Habitat moisture		Irradiance level	
	AIC-selected	simplified	AIC-selected	simplified	AIC-selected	simplified
0	39.5	38.4	45.2	41.6	63.7	59.1
1	34.3 (73.8)	34.9 (73.2)	28.0 (73.2)	29.6 (71.2)	26.1 (89.9)	29.4 (88.5)
2	16.9 (90.7)	16.9 (90.1)	17.1 (90.3)	19.8 (91.0)	6.7 (96.5)	8.0 (96.5)
3	6.8 (97.5)	7.3 (97.4)	7.1 (97.4)	6.8 (97.8)	2.4 (98.9)	2.4 (98.9)
4	2.3 (99.8)	2.4 (99.8)	2.2 (99.6)	2.0 (99.9)	1.1 (100.0)	1.1 (100.0)
5	0.2 (100.0)	0.2 (100.0)	0.4 (100.0)	0.1 (100.0)	0	0

574

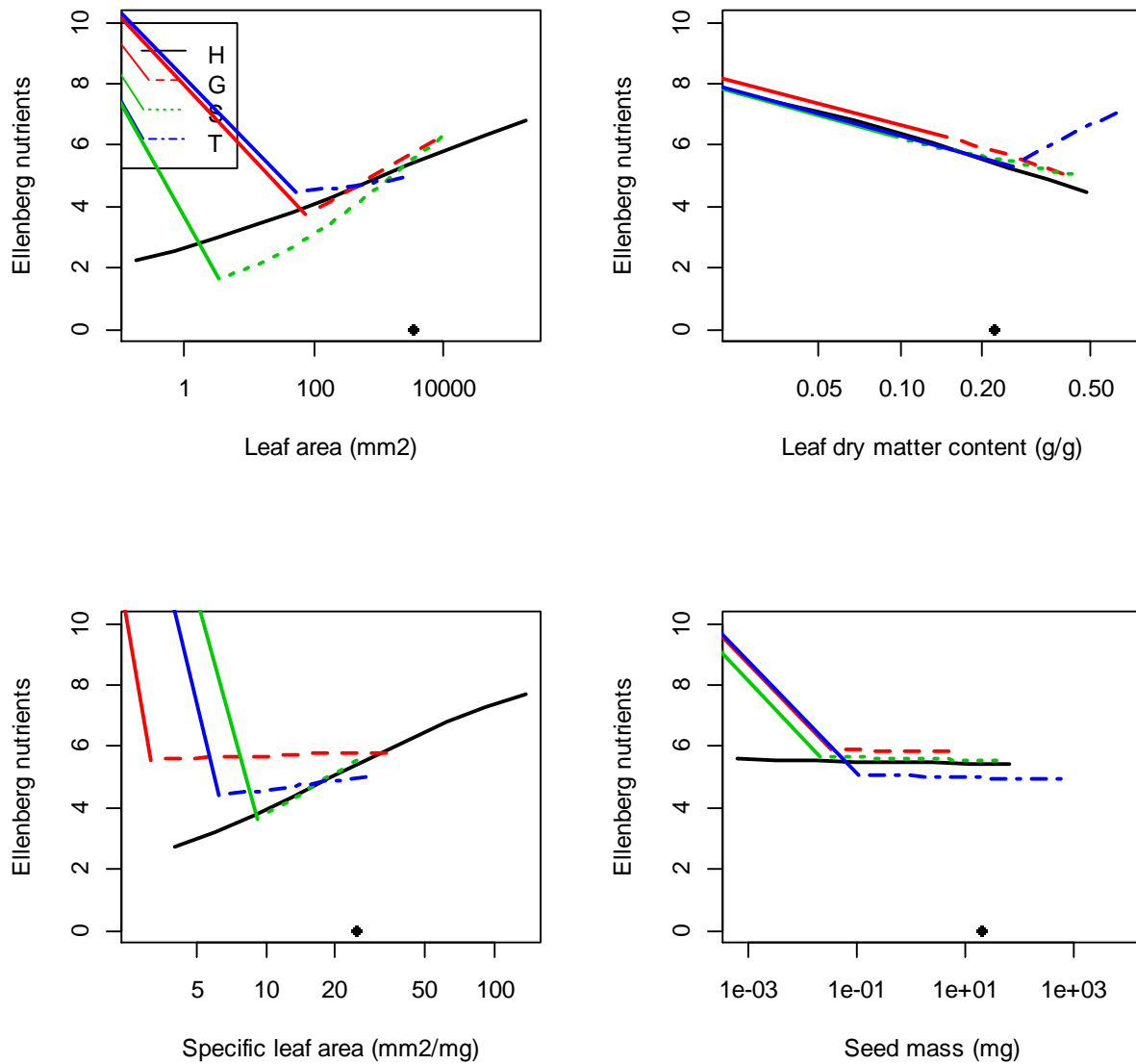
575 Figure 1. The observed Ellenberg indicator values for soil nutrient, soil moisture and irradiance plotted
576 against the mean Ellenberg indicator values predicted by four traits: specific leaf area, leaf dry matter
577 content, leaf area and seed mass. Results based on 922 (nutrients), 981 (moisture) and 988 (irradiance)
578 species.



579

580

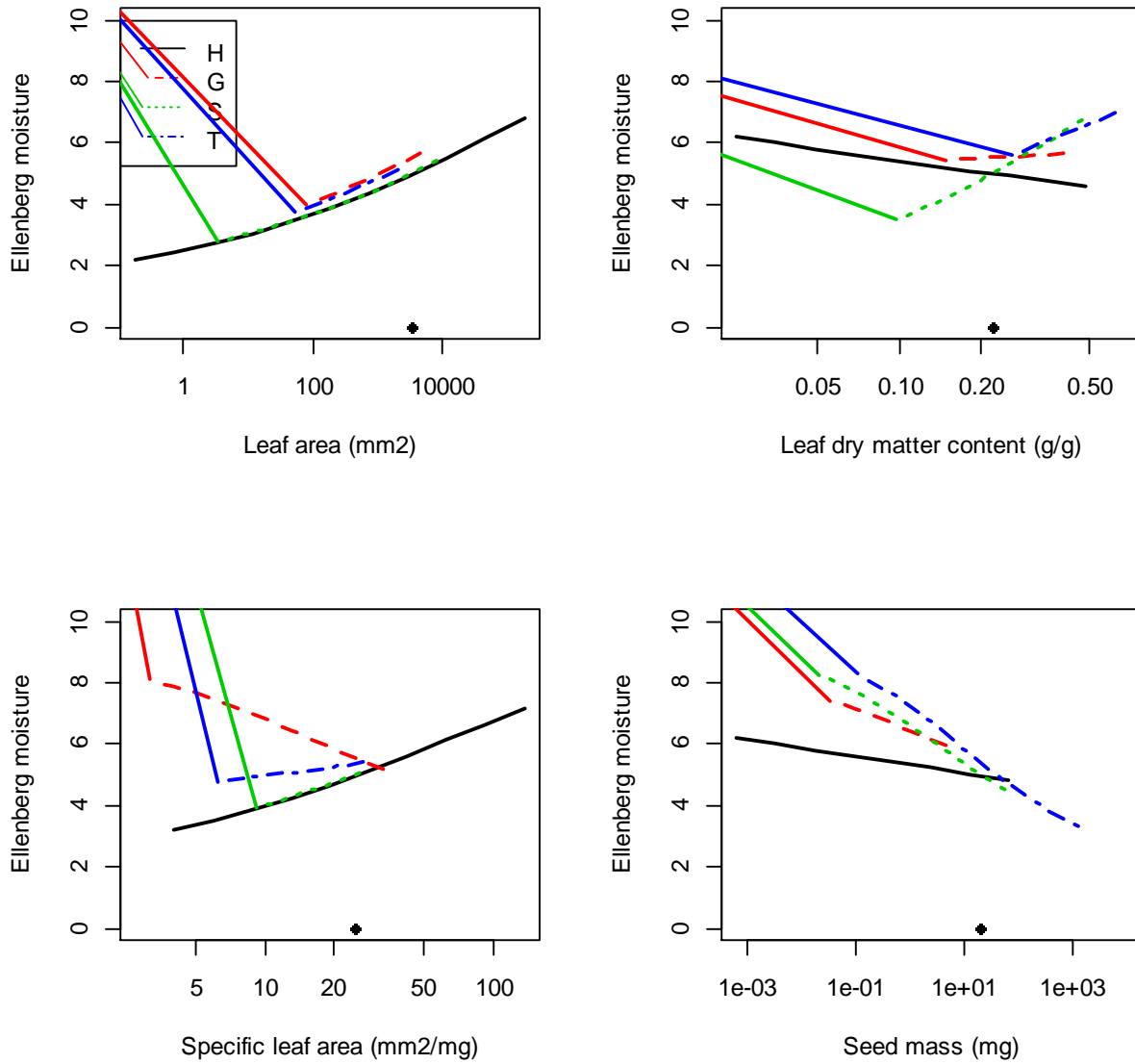
581 Figure 2. The predicted Ellenberg indicator values for soil nutrient status as a function of each plant trait.
 582 Results are shown for four plant types (H: herbs, G: graminoids, S: shrubs, T: trees). Each line shows the
 583 response for that trait when holding constant the other traits at their mean values in the data set. The
 584 mean value of each trait is shown by the dot on the x-axis..



585

586

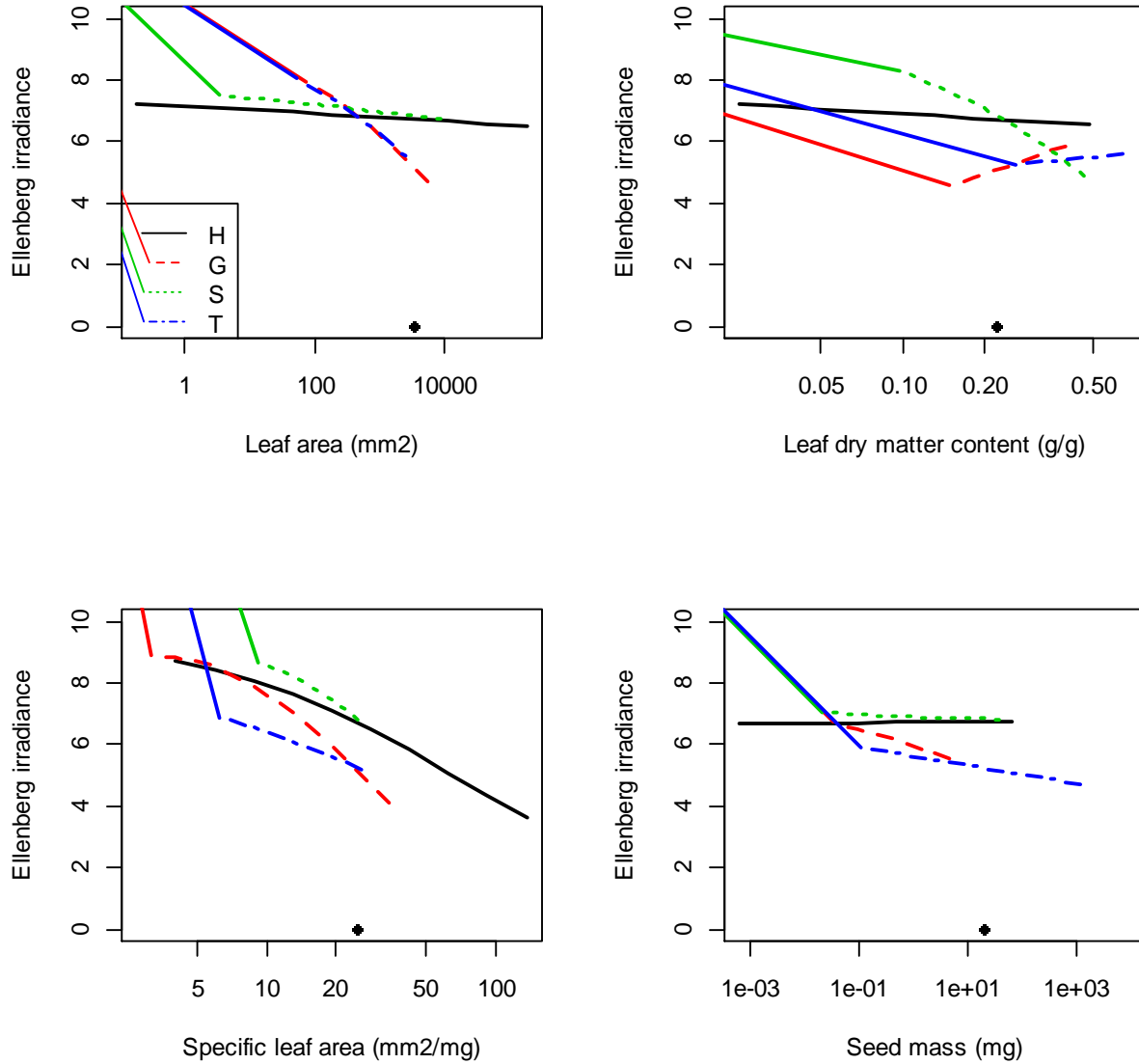
587 Figure 3. The predicted Ellenberg indicator values for soil moisture status as a function of each plant
588 trait. See Figure 1 for details.



589

590

591 Figure 4. The predicted Ellenberg indicator values for irradiance as a function of each plant trait. See
 592 Figure 1 for details.

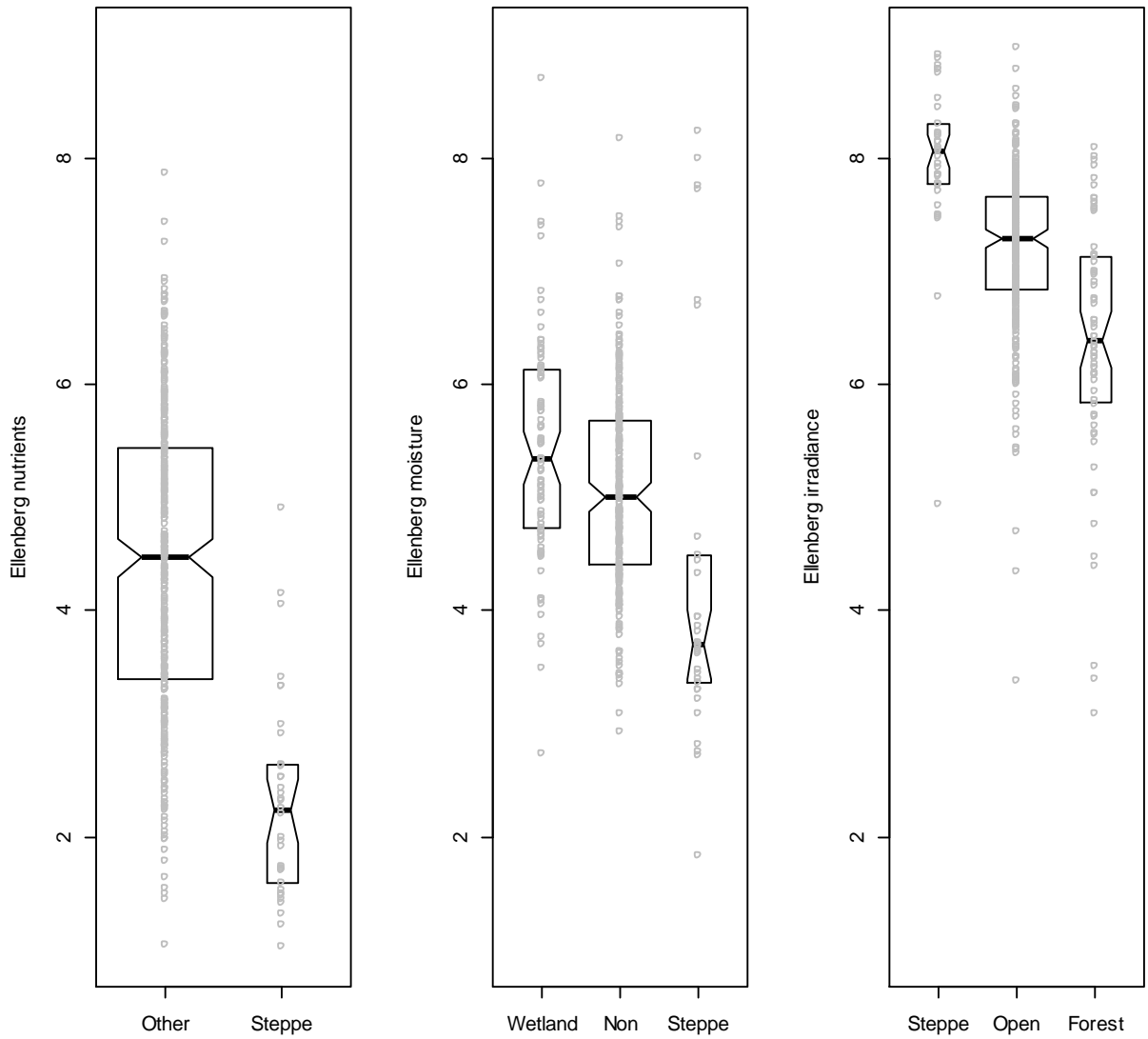


593

594

595 Figure 5. Predicted mean Ellenberg indicator values for soil nutrient and moisture status, as well as site
596 irradiance. Predictions were obtained for species that had never been classified by Ellenberg. Species
597 classified as being from the dry, infertile Moroccan steppe or not (soil nutrients), as being wetland, non-
598 wetland or Moroccan steppe species (soil moisture) and as being forest understories, open habitats or
599 from the Moroccan steppe.

600



601

602