

1 **Concepts & Synthesis**

2

3 **Favourability: concept, distinctive characteristics and potential usefulness**

4 Pelayo Acevedo^{1,2,3,4*} and Raimundo Real¹

5

6 1- Biogeography, Diversity, and Conservation Research Team, Department of Animal
7 Biology, Faculty of Sciences, University of Malaga, E-29071 Malaga, Spain.

8 2- CIBIO/UP – Centro de Investigação em Biodiversidade e Recursos Genéticos da
9 Universidade do Porto, Campus Agrario de Vairão, 4485-661 Vairão, Portugal.

10 3- Instituto de Investigación en Recursos Cinegéticos (CSIC-UCLM-JCCM), E-13071
11 Ciudad Real, Spain.

12 4- Centre de Recerca en Sanitat Animal (CRESA), UAB-IRTA, Campus de la
13 Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

14

15 *Corresponding author: pacevedo@uma.es (P. Acevedo). Tel.: +34 952132383; fax:
16 +34 952131668.

17

18 **Abstract**

19 The idea of analysing the general favourability for the occurrence of an event was
20 presented in 2006 through a mathematical function. However, even when favourability
21 has been used in species distribution modelling, the conceptual framework of this
22 function is not yet well perceived among many researchers. The present paper is
23 conceived for providing a wider and more in-depth presentation of the idea of
24 favourability; concretely we aimed to clarify both the concept and the main distinctive
25 characteristics of the favourability function, especially in relation to probability and

26 suitability, the most common outputs in species distribution modelling. As the
27 capabilities of the favourability function go beyond species distribution modelling, we
28 also illustrate its usefulness for different research disciplines for which this function
29 remains unknown. In particular, we stressed that the favourability function has potential
30 to be applied in all the cases where the probability of occurrence of an event is
31 analyzed, such as, for example, habitat-selection or epidemiological studies.

32 **Keywords:** epidemiology, favourability function, habitat selection, habitat suitability,
33 probability of occurrence, species distribution modelling.

34

35 **Brief introduction**

36 The favourability function – defined in Real et al. (2006) – assesses the variation in the
37 probability of occurrence of an event in certain conditions with respect to the overall
38 prevalence of the event. Consequently, it has potential to be applied in the cases where
39 the probability of occurrence of an event is analyzed, such as species distribution
40 modelling (Franklin 2009) or, among others, habitat-selection and epidemiological
41 studies (Manly et al. 2002; Pfeiffer et al. 2008). In addition, it can be applied to all
42 methods able to produce probability; although favourability was usually calculated from
43 probabilities yielded by logistic regressions (Hosmer and Lemeshow 2000),
44 favourability values can be derived, for example, from probabilities obtained using
45 additive or Bayesian models (Hastie and Tibshirani 1990; Bernardo and Smith 2000).
46 So far the concept and the main distinctive characteristics of favourability are not well
47 perceived among many researchers, especially for disciplines different from species
48 distribution modelling. The main aim of this study was to carry out a broader
49 presentation of the favourability concept and to illustrate the usefulness of the
50 favourability function to the scientific community.

51

52 **Defining the favourability idea and function**

53 Pierre-Simon Laplace defined probability in his first general principle about probability
54 calculation as the ratio of the number of favourable cases to the whole number of
55 possible cases (Laplace 1825, page 12). In this way, the concept of favourability was
56 implicit from the beginning in that of probability. If all cases are equally, and totally,
57 favourable – or unfavourable – then this ratio depends on the prevalence of the event. In
58 his second principle Laplace stated that different cases could differ in possibility,
59 conferring gradualness to the denominator in the probability ratio. However, it can be

60 argued that the concept of possibility is not appropriate to be given a continuous and
61 gradual value, as an event is completely possible even when it is highly unlikely, i.e.,
62 the event is completely possible if it is not completely impossible. Laplace's second
63 principle makes sense, however, if it is applied instead to the numerator of the
64 probability ratio, so pointing to a quality of each case which may be appropriately called
65 favourability and may take continuous values that can be constrained to range between
66 0 and 1. Thus, the probability of an event occurring in certain conditions combines the
67 general prevalence of the event and the local favourability for that event occurring
68 precisely in those conditions. Favourability may thus be obtained as a function of
69 probability and prevalence.

70 The favourability function was conceptually conceived in this context to assess and
71 remove the effect of prevalence on each probability value. With the favourability
72 function, output values for different events are levelled in relation to each event's
73 prevalence in the dataset. That is, a favourability value of 0.5 for an event in certain
74 locality or conditions indicates that the probability for the event's occurrence in that
75 locality or condition is the same as the overall prevalence of the event in the dataset, i.e.,
76 local conditions neither increase nor decrease the probability of occurrence with respect
77 to what could be expected according to mere prevalence, thus denoting neutral local
78 favourability. Consequently, local favourability values higher than 0.5 indicate
79 characteristics that favour the event's occurrence and values below 0.5 denote
80 detrimental conditions for the event, regardless of the event prevalence.

81 The mathematical rationale for the favourability function is presented in Real et al.
82 (2006). Basically, the favourability function may take a form similar to the logistic
83 probability in which the effect of the event's prevalence is mathematically eliminated in
84 the *logit* of a logistic regression equation. Among other forms, favourabilities (F) may

85 be directly derived from probabilities (P) yielded by any mathematical method in the
86 following way:

87
$$F = \frac{\frac{P}{(1-P)}}{\frac{n_1}{n_0} + \frac{P}{(1-P)}}$$
 being n_1 and n_0 the respective number of positive and negative

88 samples for a given event in the dataset, and, when using logistic regression, $P = \frac{e^y}{1 + e^y}$;

89 where e is the basis of the natural logarithm, and y is a regression equation of the form:

90 $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$; where α is a constant and $\beta_1, \beta_2, \dots, \beta_n$ are the

91 coefficients of the n predictor variables x_1, x_2, \dots, x_n (Tabachnick and Fidell 1996, page
92 127).

93 It must be stressed that the favourability function does not provide a probability output

94 independent of the sample prevalence, but a measure of the degree to which local

95 conditions lead to a local probability higher or lower than that expected at random,

96 being this random probability defined by the overall prevalence of the event, which in

97 turn is what must be expected if maximum entropy is assumed (Real et al. 2006). Local

98 probability depends both on the response of the dependent variable to the predictors and

99 on the overall prevalence of the event (e.g. Cramer 1999), whereas favourability values

100 depend only on the response of the dependent variable to the predictors in the study area

101 (see below). Thus, favourability is not aimed at replacing probability but at

102 complementing it, by providing, for example, a comparable measure of the response of

103 each event to the predictors for events differing in prevalence. In this way, favourability

104 may be used to detect, for example, conditions that favour in the same degree the

105 occurrence of a rare disease and a common seasonal flu, even when the probability of

106 suffering them differs due to their different prevalence. However, this concept was

107 recently misunderstood as a way to obtain the probability of occurrence when event
108 prevalence differs from 50% (Albert and Thuiller 2008).

109

110 **Sample prevalence dependence: a statistical assessment for probability,**
111 **favourability and suitability outputs**

112 To bring to light the sample prevalence dependence in the probability, favourability and
113 suitability outputs, we built a virtual species with a prevalence of 20% which was
114 designed to logistically respond to an environment defined by only one environmental
115 variable on a virtual landscape composed of 1000 units (i.e. 200 presences). From the
116 species distribution, two samples of 125 territorial units with contrasted prevalences –
117 one with 20% and another with 80% (see Figure 1) – were randomly extracted. Each
118 sample was modelled using different procedures.

119 We compared the output of the favourability function (Real et al. 2006) with those
120 resulting from probability and suitability obtained with Ecological Niche Factor
121 Analysis (ENFA; Hirzel et al. 2002) and Maximum Entropy approach (MaxEnt; Phillips
122 et al. 2006). Probabilities were obtained using logistic regression (Hosmer and
123 Lemeshow 2000), and they were included as inputs into the favourability function (Real
124 et al. 2006). ENFA was run in Biomapper 4.0 (freely available at
125 <http://www.unil.ch/biomapper/>) with the median algorithm (Hirzel et al. 2008). MaxEnt
126 version 3.1 (freely available at <http://www.cs.princeton.edu/~schapire/maxent/>) was run
127 with default parameter values and the logistic output format (Elith et al. 2011).

128 Results of all models were projected to the whole landscape (Figure 2) and outputs
129 obtained from samples with a prevalence of 20% and those with a prevalence of 80%
130 were graphically compared (Figure 3), so that outputs independent from prevalence
131 should yield a line close to the identity line. Figures 2 and 3 illustrate how the

132 favourability function was the method most independent of prevalence, since quite
133 similar results were obtained from samples with contrasted prevalences. But this did not
134 occur for probability or the suitability outputs obtained from ENFA and MaxEnt. Slight
135 mismatches observed with respect to the diagonal for the favourability function in
136 Figure 3 are due to slightly different detected responses to the variable in each randomly
137 selected sample. Our results contradicted those reported by Albert and Thuiller (2008)
138 in which favourability was suggested to be biased by sample prevalence, but they are
139 consistent with previous studies and with the conceptual framework behind
140 favourability (see Real et al. 2006).

141 The modelled response of the virtual species to the variable was the same (and correct)
142 for the probability and favourability functions, being the differences in the results only
143 due to the effect of sample prevalence on the probability outputs. Two different
144 responses of the species were obtained for ENFA and MaxEnt. ENFA was not able to
145 detect the subjacent monotonic response of the species to the environment. With both
146 samples, ENFA identified Gaussian responses (e.g. Acevedo et al. 2007) and the
147 maximum response value was obtained in both cases because in this procedure
148 suitability values are rescaled (Hirzel et al. 2002). For these reasons, two different
149 relationships were established between suitability values derived from the different
150 samples (one in each tail of the curve), but none of them was close to the identity line.

151 The results obtained for MaxEnt show that quite different responses were modelled on
152 each sample, which may be related to the fact that MaxEnt produces a number of
153 indices that are not directly related to the probability of occurrence (Royle et al. 2012).
154 Thus, with ENFA and MaxEnt the response of the species to the environment cannot be
155 segregated from the effect of sample prevalence on the suitability output. The results
156 here provided show that probability and suitability are biased in their outputs when

157 working with samples – of the same species – differing in prevalence, which is not the
158 case with favourability.

159

160 **The concept of favourability for biogeographers**

161 Many researchers working with species distribution models produce maps showing
162 continuous gradients of how environmental characteristics are appropriate – in a broad
163 sense – for a target species (Guisan and Thuiller 2005). Model's predictions can be
164 either considered as gradients or used only to classify localities as appropriate on
165 inappropriate, but the latter option limits the informative capacity of the model. Thus,
166 when models are aimed to guide conservation strategies they are more useful as
167 continuous gradients (Barbosa et al. 2010).

168 Nevertheless, the continuous model's predictions should be levelled in order to
169 determine those characteristics in the study area which actually favour the species
170 presence. That is what the favourability function does. So, using the favourability
171 function those localities with environmental conditions that favour the presence of the
172 species ($F > 0.5$) can be easily distinguished from those with detrimental characteristics
173 ($F < 0.5$) for its presence. This makes the favourability function particularly useful in
174 conservation biology, for example, to identify expansion routes of invasive species
175 (Muñoz and Real 2006; Nielsen et al. 2008), or to identify areas where a species may be
176 more vulnerable to habitat or climate changes (e.g. Guitiérrez-Illán et al. 2010).

177 The concept behind the favourability function was also raised by biogeographers
178 working with probability (Liu et al. 2005; Jiménez-Valverde and Lobo 2007) and
179 profile methods (see Hirzel et al. 2006). A rationale conceptually close to favourability
180 was used to reclassify the suitability scores obtained with the ENFA (Hirzel et al. 2002).
181 The suitability score over which the model predicts more presences than expected by

182 chance can be used as a threshold to identify the localities that actually are favourable
183 for the target species. Liu et al. (2005) and Jiménez-Valverde and Lobo (2007), for
184 example, proposed several methods to obtain the best threshold to split the localities
185 into two categories, which tend to locate the threshold near the point where probability
186 equals prevalence. These categories could appropriately be called favourable and
187 unfavourable, as they represent probabilities higher or lower than prevalence,
188 respectively. So, the determination of those conditions enhancing the probability of
189 species presence over the probability expected by chance – the concept behind the
190 favourability function – is widely considered sound in biogeography. The favourability
191 function not only provides the favourability threshold more easily ($F=0.5$) but also
192 provides information about the degree to which every locality is favourable.
193 In addition, the favourability function has other distinctive characteristics that make it
194 especially applicable in conservation biogeography and other research disciplines.

195

196 **Main distinctive characteristics of favourability values**

197 The main distinctive characteristics of the favourability function in relation to common
198 outputs in other modelling techniques (probability and/or suitability) are summarised in
199 the following five points:

200 1- Given the definition of favourability as the assessment (between 0 and 1) of the
201 variation in the probability of occurrence of an event in certain conditions with respect
202 to the overall prevalence of the event, there is only one way of obtaining favourability
203 values from probabilities and prevalences. In this aspect favourability differs from
204 suitability, as for each modelling method, suitability is an idiosyncratic way of ranking
205 local sites according to their capacity to hold the species that is not directly related to
206 probability (e.g. Guisan and Zimmermann 2000). This is why different modelling

207 techniques produce differing suitability values with the same dataset, but all ways of
208 obtaining favourability should yield the same favourability values from the same
209 dataset.

210 2- Favourability values – like probability values and unlike suitability – are
211 interpretable in absolute terms, as they indicate how local presence’s probability differs
212 from that expected by chance in the whole sample. However, suitability values, such as,
213 for example, those derived from ENFA, ensemble forecasting approaches (Araújo and
214 New 2007) or some of the outputs from MaxEnt (Phillips et al. 2006), are only relative
215 and therefore uninformative in absolute terms. For example, the suitability value
216 assigned to each focal locality in ENFA for each factor axis is based on a count of all
217 localities with species presence that lay as far or farther apart from the median than the
218 focal locality (Hirzel et al. 2002). This count is normalized in such a way that the
219 suitability index always ranges from zero to one (see Figure 3). In ensemble forecasting
220 suitability values are the result of merging, in some occasions, methods generating
221 probability with others that yield suitability scores (e.g. Thuiller et al. 2009).

222 Consequently, the suitability values obtained by these kinds of methods cannot be easily
223 interpreted, especially when comparing different models, even if they are calibrated
224 against a dataset with equal species prevalence.

225 3- Favourability values – like suitability values and unlike well calibrated probability –
226 are dependent on the extent of the study area if modifying the extent entails a
227 modification of the species prevalence. Conceptually, a locality where the probability of
228 finding a species is intermediate should be considered unfavourable for the species in
229 the context of the core of the species range, but highly favourable in the context of a
230 huge area where the species range represents a small portion. The favourability function
231 quantifies this difference of consideration of a same probability value according to the

232 differing prevalence of the species in – and normally due to the different extent of – the
233 background area. This implies that favourability (and suitability) values obtained from
234 models built in different study areas should be compared with these characteristics in
235 mind, as each favourability is relative to its own study area (Barbosa et al. 2009).

236 4- The inherent quality of the favourability function of being expressed in relation to the
237 event's prevalence in the study area enables direct comparison and combination when
238 several species are involved in the analytical design. For example, this is needed when
239 using models for multiple species as a basis for defining relevant areas for conservation
240 (Estrada et al. 2008), which cannot be built based on probability values because these
241 are higher in common than in rare species, so the values for the former would prevail
242 over those for the latter.

243 5- In addition, but closely related to point 4, favourability values – unlike probability or
244 suitability values – can be regarded as the degree of membership of the localities to the
245 fuzzy set of sites with conditions that are favourable for the species, which enables the
246 easy application of fuzzy logic operations to distribution modelling (e.g. Robertson et
247 al. 2004). Fuzzy logic operations expand the potential of the favourability function for
248 comparison between models. For example, this function and the fuzzy indices derived
249 from it were successfully used to study the biogeographical relationships in predator-
250 prey systems (Real et al. 2009) and also between native and exotic sympatric species
251 (Acevedo et al. 2010). Similarly, the transferability of models to other times, for
252 example in climate change scenarios (Real et al. 2010; Acevedo et al. 2012) or land use
253 changes (Acevedo et al. 2011), or to different resolution scales (Barbosa et al. 2010),
254 can be better assessed with the combined use of the favourability function and fuzzy
255 logic. For instance, an overall assessment of expected modification in species'
256 distribution in climate change scenarios can be obtained using fuzzy logic, since the

257 favourability forecasted for a given species in the future can be deconstructed into the
258 percentage that is expected to increase, overlap, be maintained and shift in relation to its
259 favourability in the present (Real et al. 2010). On this point, it is worth mentioning that
260 the spatial-temporal transference of models is risky and invites caution and careful
261 considerations (e.g. Jiménez-Valverde et al. 2011). There have been increasing concerns
262 about the use of correlative models for projecting species distribution into novel
263 situations such as new territories or future climate change scenarios (e.g. Sutherst and
264 Bourne 2009; Webber et al. 2011). Nevertheless, it should be noted that the concepts of
265 favourability, probability and suitability are equally applicable to mechanistic and
266 correlative modelling approaches, as they refer to the output which is produced by the
267 models, and not to the inference method used to obtain these outputs.

268

269 **The potential of the favourability function**

270 To date, applications of the favourability function are nearly restricted to species
271 distribution modelling , which is likely because the main research discipline of the
272 developers was biogeography. Taking into account the concept behind this function and
273 the distinctive characteristics of favourability values previously described, and similarly
274 to other logistic models (e.g. Keating and Cherry 2004 and references therein), the
275 potential of the favourability function in other research disciplines is high. The concept
276 of favourability is quite relevant, for instance, in habitat-selection studies for
277 determining the sampling units in which the process under study, e.g., nesting success,
278 is favoured, i.e., those sampling units with a higher probability of event occurrence than
279 expected by chance. For processes differing in prevalence favourability values provide
280 comparable measures of the response of each process to the predictors; for example,
281 with the favourability function it is possible to quantify in the same terms the degree to

282 which the local environmental characteristic are favouring bird nesting occurrence and
283 nesting success for each sampling unit (see Amici et al. 2009). In another example, Real
284 et al. (2009) used the favourability function to identify areas autoecologically
285 favourable for the rare Iberian lynx (*Lynx pardinus*) but autoecologically unfavourable
286 for its common staple prey the wild rabbit (*Oryctolagus cuniculus*), so highlighting the
287 lack of trophic resources in parts of the potential range for a critically endangered
288 species. This would be unattainable with probabilities, as the very common, and
289 prevalent, rabbit tend to yield higher values of probability of occurrence than the scarce
290 lynx, even in localities where rabbit densities are unable to support lynx populations.
291 The concept of favourability and its distinctive characteristics are also promising in
292 epidemiology. Epidemiological studies in wildlife try to identify risk factors that
293 increase the frequency of pathogens (e.g. Vicente et al. 2007) and to create risk maps in
294 which the probability of their transmission is shown (e.g. Rochlin et al. 2011). Including
295 the concept of favourability in these studies entails two main advantages. First, those
296 populations (or individuals, it depends on the sampling unit used in the study) in which
297 the probability of presence of the pathogen is higher than expected by chance ($F > 0.5$) in
298 the study area can be identified. These are key populations for disease control and
299 monitoring (Mörner et al. 2002). Similarly, those values of a given risk factor over
300 which the probability for the presence of a pathologic condition is higher than expected
301 by chance can also be identified. For example, Fernández et al. (2000) studied the
302 relationships between coronary artery anomalies and aortic valve morphology obtaining
303 that the probability of occurrence of anomalous coronary artery patterns increases
304 continuously according to the degree of deviation of the aortic valve from its normal
305 (tricuspid) design according, for example, to the following logit expression: $y = -2.0976$
306 $+ 0.3136 * \text{group}$ (where group referred to six groups of valve conditions into which the

307 continuous spectrum of aortic valve morphology was divided, from 0=tricuspid to
308 5=bicuspid). By including the favourability concept in this study, the authors could
309 have determined over which aortic valve morphotype (from 0 to 5) the probability of
310 occurrence of the anomalous coronary pattern was higher than expected by chance, and
311 therefore, the anomalies were being promoted. So, given the expression previously
312 reported and considering that 220 out of 968 of the coronary artery patterns were
313 anomalous, a favourability value higher than 0.5 is obtained for valve morphotype value
314 higher than 2.7, so these are the values that actually favour the anomalous coronary
315 pattern.

316 Secondly, favourability is also a promising function for biogeography of diseases
317 where interactions among – hosts and vectors – species differing in prevalence are
318 relevant (Peterson 2008) and where time series are usually available (e.g. Boadella et al.
319 in press). As previously stated, the use of the favourability function and fuzzy logic
320 allows direct comparisons and/or combinations between more than one model (host,
321 vector and pathogen), which enables a more complete assessment of the distribution of
322 the disease transmission risks (see Estrada-Peña et al. 2008) by obtaining reliable multi-
323 host, multi-pathogen and/or multi-scenario risk maps. In this context, Boadella et al. (in
324 press) analyzed the factors associated to the detection of a group of parasites –
325 *Trichinella* spp. – infecting wild boar (*Sus scrofa*). The inclusion of the idea of
326 favourability in this study (first time in spatial epidemiology) was needed to combine
327 the risks obtained for each of the 12 years included in the study in order to obtain two
328 proxies of the risk for *Trichinella* spp. infection for the study period. One index was
329 defined to identify areas where the conditions for *Trichinella* spp. infection were
330 favourable during the study period (endemic areas for the parasites), and another was
331 designed to determine the global distribution of these parasites during the study period.

332 So, the combined use of the favourability function and fuzzy logic operations enabled
333 a more-in-depth assessment of the risks for a given parasite group in a multi-scenario
334 context.

335

336 **Concluding remarks**

337 The main aim of this study was to carry out a broad presentation of the favourability
338 concept and the favourability function to the scientific community. In addition to the
339 studies in conservation biogeography, here we highlighted the usefulness of this
340 function in two other disciplines (habitat-selection and epidemiology). We think that its
341 capabilities go beyond these examples, and that the examination of the concept and the
342 exploration of its usefulness for other disciplines will prove to be helpful in all cases
343 where the probability of occurrence of an event is analyzed.

344

345 **Acknowledgements**

346 The authors acknowledge funding from Spanish Plan Nacional de Investigación and
347 FEDER CGL2009-11316/BOS. P. A. was supported by the Vicerrectorado de
348 Investigación of the University of Malaga and currently by a Beatriu de Pinós
349 fellowship funded by Comissionat per a Universitats i Recerca del Departament
350 d'Innovació, Universitats i Empresa, of the Generalitat de Catalunya, and the COFUND
351 Programme - Marie Curie Actions under 7th Marc Programme of the European
352 Community.

353

354 **References**

355 Acevedo P, Cassinello J, Hortal J, Gortázar C (2007) Invasive exotic aoudad
356 (*Ammotragus lervia*) as a major threat to native Iberian ibex (*Capra pyrenaica*): a
357 habitat suitability model approach. *Divers Distrib* 13:587-597.

358 Acevedo P, Farfán MA, Márquez AL, Delibes-Mateos M, Real R, Vargas JM (2011)
359 Past, present and future of wild ungulates in relation to changes in land use.
360 *Landscape Ecol* 26:19-31.

361 Acevedo P, Jiménez-Valverde A, Melo-Ferreira J, Real R, Alves PC (2012) Parapatric
362 species and the implications for climate change studies: a case study on hares in
363 Europe. *Global Change Biol* doi: 10.1111/j.1365-2486.2012.02655.x

364 Acevedo P, Ward AI, Real R, Smith GC (2010) Assessing the biogeographical
365 relationships of ecologically related species using favourability functions: a case
366 study on British deer. *Divers Distrib* 16:515-528.

367 Albert CH, Thuiller W (2008) Favourability functions versus probability of presence:
368 advantages and misuses. *Ecography* 31:417-422.

369 Amici A, Pelorosso R, Serrani F, Boccia L (2009) A nesting site suitability model for
370 rock partridge (*Alectoris graeca*) in the Apennine Mountains using logistic
371 regression. *Ital J Anim Sci* 8:751-753.

372 Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol*
373 *Evol* 22:42-47.

374 Barbosa AM, Real R, Vargas JM (2009) Transferability of environmental favourability
375 models in geographic space: The case of the Iberian desman (*Galemys*
376 *pyrenaicus*) in Portugal and Spain. *Ecol Model* 220:747-754.

377 Barbosa AM, Real R, Vargas JM (2010) Use of Coarse-Resolution Models of Species'
378 Distributions to Guide Local Conservation Inferences. *Conserv Biol* 24: 1378-
379 1387.

380 Bernardo JM, Smith AFM (2000) Bayesian Theory. Wiley Series in Probability and
381 Statistics. John Wiley & Sons, Hoboken.

382 Boadella M, Barasona JA, Pozio E, Montoro V, Vicente J, Gortázar C, Acevedo P (in
383 press) Spatial-temporal trend and risk factors for *Trichinella* sp. infection in wild
384 boar (*Sus scrofa*) populations of central Spain: a long-term study. Int J Parasitol
385 Cramer JS (1999) Predictive performance of binary logit model in unbalanced samples.
386 J R Stat Soc D 48:85-94.

387 Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical
388 explanation of MaxEnt for ecologist. Divers Distrib 17:43-57.

389 Estrada A, Real R, Vargas JM (2008) Using crisp and fuzzy modelling to identify
390 favourability hotspots useful to perform gap analysis. Biodivers Conserv 17:857-
391 871.

392 Estrada-Peña A, Acevedo P, Ruiz-Fons F, Gortázar C, de la Fuente J (2008) Evidence
393 of the Importance of Host Habitat Use in Predicting the Dilution Effect of Wild
394 Boar for Deer Exposure to *Anaplasma* spp. PLoS ONE 3(8): e2999.
395 doi:10.1371/journal.pone.0002999.

396 Fernández MC, Durán AC, Real R, López D, Fernández B, de Andrés AV, Arqué JM,
397 Gallego A, Sans-Coma V (2000) Coronary artery anomalies and aortic valve
398 morphology in the Syrian hamster. Lab Ani-UK 34:145-154.

399 Franklin J (2009) Mapping species distributions. Spatial inference and prediction.
400 Cambridge University Press, Cambridge.

401 Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple
402 habitat models. Ecol Lett 8:993-1009.

403 Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology.
404 Ecol Model 135:147-186.

405 Gutiérrez-Illán J, Gutiérrez D, Wilson RW (2010) The contributions of topoclimate and
406 land cover to species distributions and abundance: fine-resolution tests for a
407 mountain butterfly fauna. *Global Ecol Biogeogr* 19:15-173.

408 Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman & Hall/CRC

409 Hirzel AH, Braunisch V, Le Lay G, Hausser J, Perrin N (2008) BIOMAPPER 4.0.
410 Laboratoy of Conservation Biology, Department of Ecology and Evolution,
411 University of Lausanne, Lausanne, Switzerland. Available from
412 <http://www.unil.ch/biomapper>.

413 Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis:
414 How to compute habitat-suitability maps without absence data? *Ecology* 83:2027-
415 2036.

416 Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of
417 habitat suitability models to predict species presences. *Ecol Model* 199:142-152.

418 Hosmer DW, Lemeshow S (2000) *Applied Logistic Regression*. Wiley Interscience,
419 New York.

420 Jiménez-Valverde A, Lobo JM (2007) Threshold criteria for conversion of probability
421 of species presence to either- or presence-absence. *Acta Oecol* 31:361-369.

422 Jiménez-Valverde A, Peterson AT, Soberón J, Overton J, Aragón P, Lobo J M (2011)
423 Use of niche models in invasive species risk assessments. *Biol Invasions* 13:2785-
424 2797.

425 Keating KA, Cherry S (2004) Use and interpretations of logistic regression in habitat-
426 selection studies. *J Wildlife Manage* 68:774-789.

427 Laplace PS (1825) *Essai philosophique sur les probabilités*. Bachelier, Paris.

428 Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in
429 the prediction of species distributions. *Ecography* 28:385-393.

430 Manly BFJ, McDonald LL, Thomas DL, McDonald TL, Erickson W (2002) Resource
431 selection by animals: statistical design and analysis for field studies. Second
432 edition. Kluwer Press, New York.

433 Mörner T, Obendorf D, Artois M, Woodford M (2002) Surveillance and monitoring of
434 wildlife diseases. *Rev Sci Tech Off Int Epiz* 21:67-76.

435 Muñoz AR, Real R (2006) Assessing the potential range expansion of the exotic monk
436 parakeet in Spain. *Divers Distrib* 12:656-665.

437 Nielsen C, Hartvig P, Kollmann J (2008) Predicting the distribution of the invasive alien
438 *Heracleum mantegazzianum* at two different spatial scales. *Divers Distrib* 14:307-
439 317.

440 Peterson AT (2008) Biogeography of diseases: a framework for analysis.
441 *Naturwissenschaften* 95:483-491.

442 Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers D, Clements A (2008) Spatial
443 analysis in epidemiology. Oxford University Press, New York.

444 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species
445 geographic distributions. *Ecol Model* 190:231-259.

446 Real R, Barbosa AM, Rodríguez A, García FJ, Vargas JM, Palomo LJ, Delibes M
447 (2009) Conservation biogeography of ecologically interacting species: the case of
448 the Iberian lynx and the European Rabbit. *Divers Distrib* 15:390-400.

449 Real R, Barbosa AM, Vargas JM (2006) Obtaining environmental favourability
450 functions from logistic regression. *Environ Ecol Stat* 13:237-245.

451 Real R, Márquez AL, Olivero J, Estrada A (2010) Are species distribution models in
452 climate warming scenarios useful for informing emission policy planning? An
453 uncertainty assessment using fuzzy logic. *Ecography* 33:304-314.

454 Robertson MP, Villet MH, Palmer AR (2004) A fuzzy classification technique for
455 predicting species' distributions: applications using invasive alien plants and
456 indigenous insects. *Divers Distrib* 10:461-474.

457 Rochlin I, Turbow D, Gomez F, Ninivaggi DV, Campbell SR (2011) Predictive
458 Mapping of Human Risk for West Nile Virus (WNV) Based on Environmental
459 and Socioeconomic Factors. *PLoS ONE* 6(8): e23280.doi:10.1371/journal.pone.
460 0023280.

461 Royle JA, Chandler RB, Yackulic C, Nichols JD (2012) Likelihood analysis of species
462 occurrence probability from presence-only data for modelling species
463 distributions. *Methods in Ecology & Evolution* doi: 10.1111/j.2041-
464 210X.2011.00182.x

465 Sutherst RW, Bourne AS (2009) Modelling non-equilibrium distributions of invasive
466 species: a tale of two modelling paradigms. *Biol Invasions* 11:1231-1237.

467 Tabachnick BG, Fidell LS (1996) *Using multivariate analysis*, 3rd edn. HarperCollins
468 College Publishers, Northridge, California.

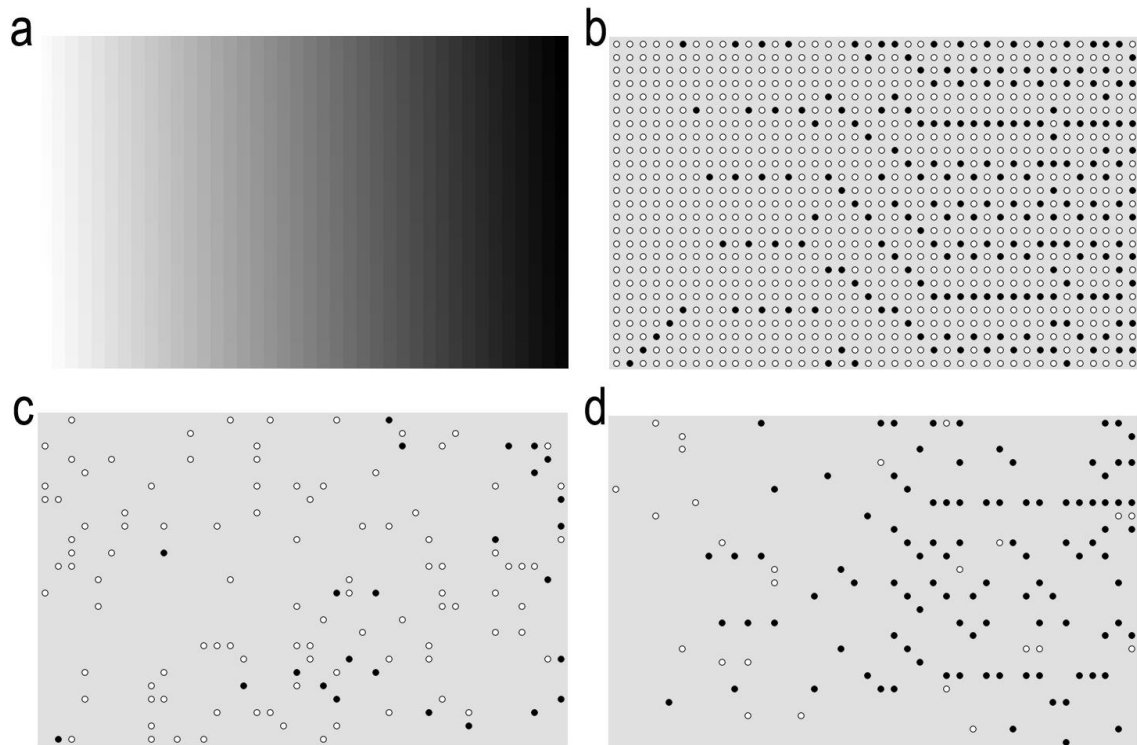
469 Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD - a platform for
470 ensemble forecasting of species distributions. *Ecography* 32:369-373.

471 Vicente J, Hofle U, Garrido JM, Fernandez-de-Mera IG, Acevedo P, Juste R, Barral M,
472 Gortázar C (2007) Risk factors associated with the prevalence of tuberculosis-like
473 lesions in fenced wild boar and red deer in south central Spain. *Vet Res* 38:451-
474 464.

475 Webber BL, Yates CJ, Le Maitre DC, Scott JK, Kriticos DJ, Ota N, McNeill A, Le
476 Roux JJ, Midgley GF (2011) Modelling horses for novel climate courses: insights
477 from projecting potential distributions of native and alien Australian acacias with
478 correlative and mechanistic models. *Divers Distrib* 17:978-1000.

479

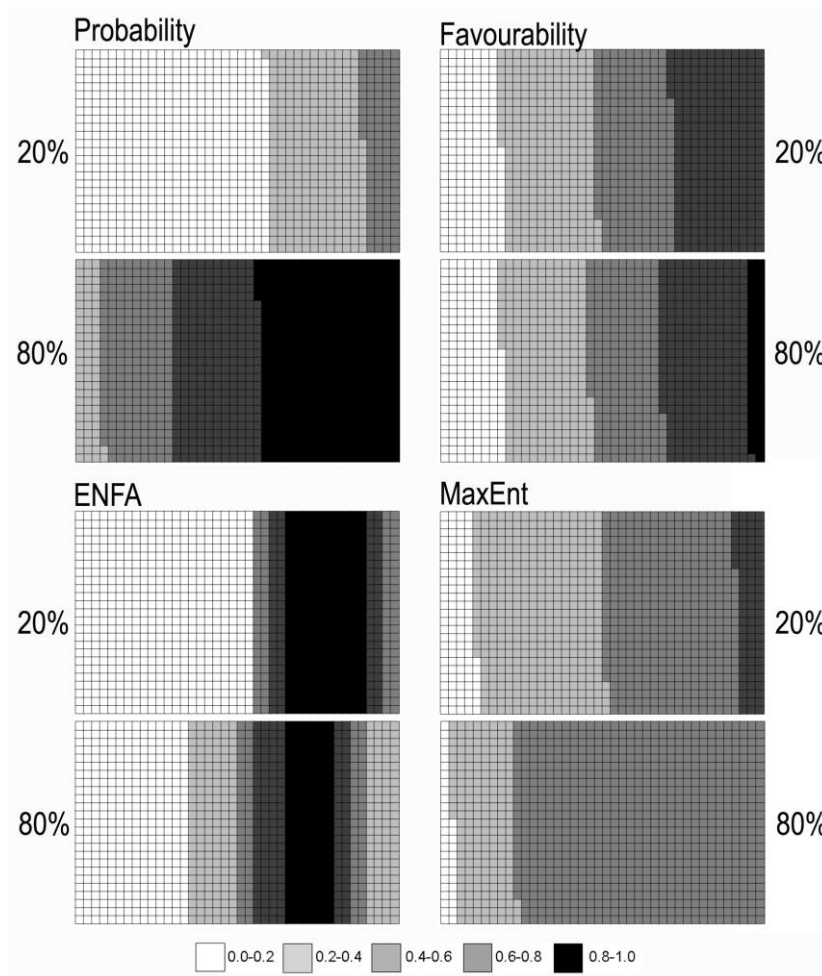
480 **Figure 1.** Virtual landscape composed of 1000 units with (a) an environmental variable
481 ranging from 0 (white) to 1000 (black); (b) a virtual species distribution (black circles
482 show presences and white ones absences); and random samples of the species with
483 prevalence of 20% (c) or 80% (d).
484



485

486

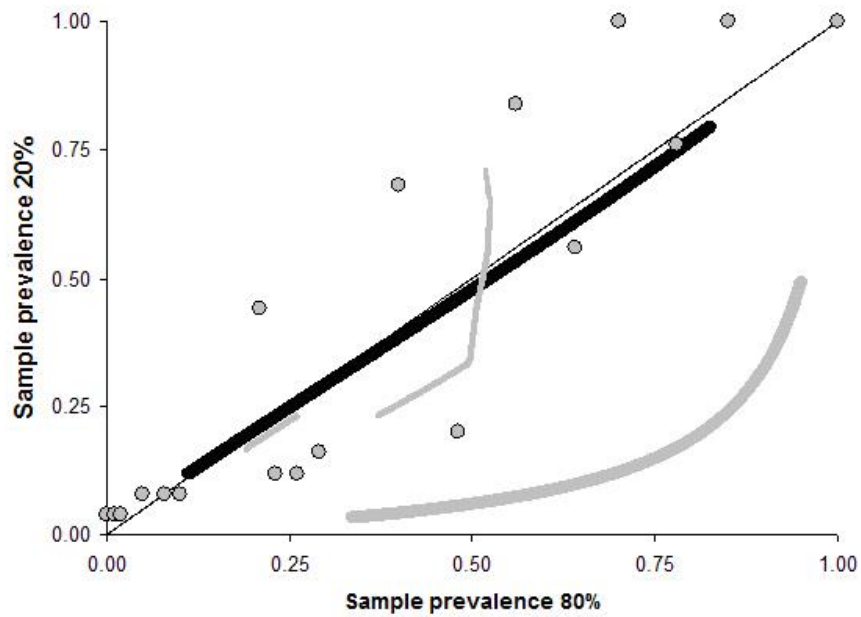
487 **Figure 2.** Predictions obtained for each sample (20% or 80%) and modelling
488 procedure (probability, favourability, and suitability from ENFA and MaxEnt).
489



490

491

492 **Figure 3.** Comparison between outputs of models developed from a sample with
493 prevalence of 20% against others from a sample with 80%. Lines are representing
494 outputs of favourability (black-thick), probability (grey-thick), and MaxEnt (grey
495 thin). Results from ENFA are represented with grey circles. The black-thin line
496 represents de identity.
497



498

499