

1 **Species distribution models predict range expansion better than chance but not**  
2 **better than a simple dispersal model**

3  
4  
5 3

6  
7 4 Marta Rodríguez-Rey<sup>1</sup>, Alberto Jiménez-Valverde<sup>2</sup>, Pelayo Acevedo<sup>2,3,4\*</sup>  
8  
9 5

10  
11  
12 6 <sup>1</sup> Area of Zoology, Department of Environmental Sciences, University of Castilla-La  
13  
14 Mancha, E-45071 Toledo, Spain.  
15 7

16  
17 8 <sup>2</sup> Biogeography, Diversity, and Conservation Research Team, Faculty of Sciences,  
18  
19 University of Málaga, E-29071 Málaga, Spain.  
20 9

21  
22 10 <sup>3</sup> Instituto de Investigación en Recursos Cinegéticos (CSIC-UCLM-JCCM), E-13071  
23  
24 Ciudad Real, Spain.  
25 11

26  
27 12 <sup>4</sup> CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio  
28  
29 Laboratório Associado. Universidade do Porto, 4485-661 Vairão, Portugal.  
30 13

31  
32 14 \* Corresponding author: Dr. Pelayo Acevedo, Tel: (351) 252 660411, Fax: (351) 252  
33  
34 15 661780, E-mail: [pelayo.acevedo@gmail.com](mailto:pelayo.acevedo@gmail.com)  
35

36  
37 16

38  
39 17 **Running title:** Predicting species range expansion  
40

41  
42 18  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 **Abstract**

2 The evaluation of species distribution models (SDMs) is a crucial step; usually, a  
3 random subsample of data is used to test prediction capacity. This procedure, called  
4 cross-validation, has been recently shown to overestimate SDMs performance due to  
5 spatial autocorrelation. In the case of expanding species, there exists the possibility to  
6 test the predictions with non-random geographically structured data, i.e., a new data set  
7 which corresponds to the last occupied localities. The aim of this study was to evaluate  
8 the capacity of SDMs to predict the range expansion pattern of six free-living deer  
9 species in Great Britain and to assess whether SDMs perform better than a simple  
10 dispersal model - a null model that assumes no environmental control in the expansion  
11 process. Distribution data for the species prior to 1972 were used to train the SDMs  
12 (ENFA, MAXENT, logistic regression and an ensemble model) in order to obtain  
13 suitability maps. Additionally, the geographical distance to the localities occupied in  
14 1972 was considered a proxy of the probability that a certain locality has to be occupied  
15 during an expansion process considering only dispersal (GD model). Subsequently, we  
16 analyzed whether the species increased their ranges between 1972 and 2006 according  
17 to the estimated suitability patterns and whether or not SDMs predictions outperformed  
18 GD predictions. SDMs showed a high discrimination capacity in the training data, with  
19 the ensemble models performing the best and ENFA models the worst. SDMs  
20 predictions also worked better than chance in classifying new occupied localities,  
21 although differences among techniques disappeared and the predictions showed no  
22 difference with respect to GD. Spatial autocorrelation of both the environmental  
23 predictors and the expansion process may explain these results which illustrate that GD  
24 is a much more parsimonious model than any of the SDMs and may thus be preferable

1 both for prediction and explanation. Overestimation of SDMs performance and  
2 usefulness may be a common fact.

3 **Key-words:** autocorrelation, deer species, geographic distance, predictive performance,  
4 range expansion, species distribution models

5

## 1. Introduction

The use of species distribution modelling (SDM) has grown exponentially in the last two decades and has shown its potential in the fields of biodiversity conservation and ecosystem management (Franklin, 2009; Peterson et al., 2011). It is usually implicitly made the assumption that species distributions are in equilibrium with the environment (Franklin, 2009). However, in nature, equilibrium is the exception rather than the rule (Gaston, 2009). Paradoxically, it is under such circumstances when species distribution models (SDMs) are particularly needed. For instance, species in the first stages of an invasion process is a clear example of such a scenario (Peterson, 2003). Explicitly recognizing non-equilibrium as a probable working scenario is important because it has methodological implications in SDM (Jiménez-Valverde et al., 2011).

At present, one of the greatest challenges in SDM is model evaluation (Vaughan and Ormerod, 2005; Lobo et al., 2008; Hijmans, 2012; Jiménez-Valverde, 2012; Jiménez-Valverde et al., 2013; Smith 2013). Discrimination capacity is the property that obtains most of the attention, and it is generally accepted that model testing should be performed on data that have not been used in the training step; otherwise, model performance would be overestimated. To get this independent testing data, modellers usually perform the so-called cross-validation, i.e., they randomly divide the data into a training set and a validation set (Fielding and Bell, 1997). However, Hijmans (2012) compared the predictive performance (AUC values) of two SDM techniques with that of a purely distance-based method (the null model) and showed that because training and testing presences are closer to each other than training presences and testing absences, cross-validation still overestimated SDMs discrimination capacity.

When working with species that are in the process of expanding their geographic ranges, model evaluation can be performed using a new data set from the most recent

1 time period (Araújo et al., 2005), which will correspond to the observed area of  
2 expansion instead of a random subsample of the data (Jiménez-Valverde et al. 2011).  
3 Usually, this new testing data set will be strongly spatially structured in some way. To  
4 avoid overestimating the distribution of the species, it has been suggested that spatial  
5 and environmental predictors should be considered together when modelling species in  
6 disequilibrium (De Marco et al., 2008; Sullivan et al., 2012). Including spatial variables  
7 in SDMs is desirable when the aim is to predict the most vulnerable localities that  
8 would be occupied in the short term during an expansion course (De Marco et al.,  
9 2008). Explicitly including the spatial variables in the modelling framework helps to  
10 maintain range cohesion (De Marco et al., 2008), and may be a way to account for  
11 spatially structured non-environmental factors, such as dispersal behaviour (Sullivan et  
12 al., 2012). These factors may significantly affect species distribution, especially in the  
13 initial phases of a range expansion process (Muñoz and Real, 2006).

14 The main objective of this study was to compare the predictive performance of four  
15 classic SDM techniques and a method that accounted for dispersal alone as a null model  
16 that assumes no environmental control in the expansion process. Using data about the  
17 recent range expansion of six species of deer in Britain, we attempted to answer the  
18 following two questions: (1) do SDMs provide significantly better-than-chance  
19 predictions of the species' range expansions? and (2) do SDMs perform better than the  
20 dispersal model?

## 22 **2. Materials and methods**

### 23 *2.1 The species*

24 There are six species of deer living wild in Britain: two are native (red deer *Cervus*  
25 *elaphus* and roe deer *Capreolus capreolus*); one is naturalized (introduced by the

1 Normans around 1000 years ago; fallow deer *Dama dama*); and three are non-native  
2 species introduced between 50 and 150 years ago (sika deer *Cervus nippon*, Reeves'  
3 Muntjac *Muntiacus reevesi* and Chinese water deer *Hydropotes inermis*). It has been  
4 estimated that the six species have expanded their ranges in Britain between 1972 and  
5 2002 (Ward, 2005), a tendency that is still occurring (Ward et al., 2008), and seems to  
6 be happening throughout Europe (Apollonio et al., 2010).  
7 Species distribution data refer to a 10 km × 10 km grid superimposed on a map of  
8 Britain comprising 2800 grid squares. For modelling purposes, the study area was  
9 restricted to 2283 grid squares to avoid potential bias in modelling arising from  
10 including those smaller than 14 ha (coastline). Data on deer species distribution (Fig. 1)  
11 were obtained from Ward (2005) and supplemented with data from Acevedo et al.  
12 (2010). The idea was to replicate a common modelling exercise in which data for a  
13 species in disequilibrium are modelled and the geographic projection is interpreted as a  
14 map of potential ways for future colonization. To do so, the data from 1972 were used  
15 to train the models and to obtain suitability maps for the six species in Great Britain.  
16 Subsequently, we analyzed whether the species increased their ranges between 1972 and  
17 2006 according to the estimated suitability surfaces, i.e., whether they expanded their  
18 ranges occupying preferably those localities with higher suitability values (as estimated  
19 using the data from 1972).

## 21 *2.2 Predictors and modelling techniques*

22 Twenty-four environmental predictors, grouped into two main factors (climate and  
23 topography), were chosen on the basis of their potential predictive power (Table 1).  
24 Although land use variables are usually taken into account when modelling the  
25 distribution of wild ungulates (e.g. Acevedo et al., 2010, 2011), they were not

1 considered in this study because, to the best of our knowledge, land use information was  
2 unavailable for the training period. Two models were run with each SDM technique (see  
3 below); one using only the environmental variables as predictors and another one which  
4 also included latitude and longitude (spatial factor; Table 1) to account for spatially  
5 structured non-environmental factors (De Marco et al., 2008; Sullivan et al., 2012).

6 Four distinct SDM techniques selected to represent different levels of model complexity  
7 and data requirements (see Jiménez-Valverde et al., 2011) were used to model the  
8 occurrence of the species (see details in Appendix A): Ecological Niche Factor Analysis  
9 (ENFA; Hirzel et al., 2002), MAXENT (Phillips et al., 2006; Phillips and Dubík, 2008),  
10 logistic regression (GLM; Hosmer and Lemeshow, 2000) and an ensemble model (EM;  
11 Thuiller et al., 2009) of four techniques (generalized linear models, multivariate  
12 adaptive regression splines, generalized boosted models and random forests).

### 13 14 *2.3 The dispersal model*

15 For each species, the geographical distance from each non-occupied locality in 1972 to  
16 the nearest occupied locality in the same period was calculated. This distance was  
17 considered a simple proxy of the probability that a certain location has to be occupied in  
18 an expansion process, i.e., it is more probable that a species will disperse to those  
19 localities that are closer to the species range limit. Geographic distances were  
20 standardized between 0 and 1 and their inverse was calculated such that a value equal to  
21 1 corresponded to those localities having the highest probability of occupancy. In this  
22 model (GD herein), a grid cell with one occupied neighbour has the same value as one  
23 with 2 or more neighbours. Also, the shortest distance between two points in geographic  
24 space does not necessarily correspond to the shortest distance a deer would have to

1 travel from one point to the other. These two complexities were deliberately avoided to  
2 make GD as simple as possible.

#### 3 4 5 6 7 2.4 Assessment of the predictive capacity and differences between techniques

8  
9  
10 5 To assess whether the species tended to occupy the localities with higher suitability  
11 values (estimated by the models trained with the data from 1972), the probability that an  
12 occupied locality chosen at random has of showing a higher estimated suitability value  
13 than an unoccupied locality chosen at random was calculated. Therefore, the area under  
14 the ROC curve (AUC) was computed to assess the predictive capacity of the models.

15  
16  
17  
18  
19  
20  
21  
22 10 The predictive performance of each model was assessed on two data sets named T1 and  
23 T2. T1 consisted of the same data used to parameterize the models, i.e., the localities  
24 occupied up to 1972 and the unoccupied ones (note that, in this case, no evaluation  
25 exists for GD). T2 provided the probability of interest, and consisted of the new  
26 localities occupied between 1972 and 2006 and the ones that were still unoccupied. We  
27 assessed differences between techniques in relation to their predictive performance  
28 using general linear mixed models (GLMMX; Zuur et al., 2009). Two GLMMX were  
29 carried out, one for each evaluation data set. The AUC was the dependent variable;  
30 “species” was included as a random factor and “technique” as a fixed factor. In addition,  
31 we included another fixed factor (“approach”) to account for differences between  
32 models with and without spatial variables. Post-hoc Tukey's tests were used to assess  
33 differences between pairs of techniques.

### 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 23 **3. Results**

54  
55  
56 24 In general, all the techniques yielded a similar spatial pattern; the greatest differences  
57 were those between ENFA and GD and the remaining methods (Appendix B).



1 Discrimination was always better than chance ( $AUC > 0.5$ ) in T1; mean AUC values  
2 were higher than 0.8 (mean  $\pm$  SD:  $0.890 \pm 0.085$ ; min-max: 0.655-1.000). There were  
3 statistically significant differences in predictive capacity between the techniques but not  
4 between models with and without spatial variables (technique:  $F_{3,38}=59.925$ ,  $P < 0.01$ ;  
5 approach:  $F_{1,38}=0.424$ ,  $P=0.519$ ; random factor:  $F_{5,38}=8.308$ ,  $P < 0.01$ ). Three groups  
6 were established (Tukey's test; from highest to lowest discrimination power): i) EM, ii)  
7 MAXENT and GLM, iii) ENFA (see Fig. 2). Differences among techniques were  
8 consistent in the two approaches and across most part of the ROC space (Appendix C).  
9 Discrimination capacity in T2 decreased compared to T1, although it was still better  
10 than chance ( $AUC > 0.5$ ) for most of the techniques (mean AUC  $\pm$  SD:  $0.760 \pm 0.116$ ;  
11 min-max: 0.492-0.929). There were statistically significant differences in discrimination  
12 capacity between the techniques but not between approaches (technique:  $F_{4,40}=7.820$ ,  
13  $P < 0.01$ ; approach:  $F_{1,43}=0.611$ ,  $P=0.439$ ; random factor:  $F_{5,43}=44.521$ ,  $P < 0.01$ ). Two  
14 groups could be established (Tukey's test; from highest to lowest discrimination power):  
15 i) GD, EM, MAXENT and GLM, ii) ENFA (see Fig. 2). Differences among techniques  
16 were consistent in the two approaches and across most part of the ROC space  
17 (Appendix C).

#### 18 19 **4. Discussion**

20 In general, deer species distributions were well estimated by the four SDM techniques.  
21 The results in T1 are consistent with the current state of knowledge: EM performed  
22 better than single techniques (Marmion et al., 2009); ENFA showed the lowest  
23 discrimination capacity (Tsoar et al., 2007); and MAXENT and GLM performed similarly  
24 (Wisiz et al., 2008). Given that we were able to account for the distribution of the  
25 species with relatively high accuracy using SDMs, we could expect some degree of

1 predictive capacity in T2; the localities unoccupied in 1972 with higher suitability  
2 values would be occupied with a higher probability than the localities with lower values  
3 (e.g. Cassinello et al., 2006; Muñoz and Real, 2006; Wilson et al., 2007; Gassó et al.,  
4 2012). Although - as expected - discrimination capacity decreased with respect to T1,  
5 the models still predicted the new occupied localities better than chance (with the  
6 exception of ENFA). As a preliminary and likely hasty conclusion, it could be said that  
7 SDM was a useful tool to forecast the range expansion for the focus species. However,  
8 three results are especially relevant: first, the inclusion of spatial variables in SDMs did  
9 not significantly improve the predictions; second, the differences in performance  
10 between SDMs that predicted better than chance in T2 disappeared; third, there were not  
11 significant differences in performance between SDMs and GD. On the one hand, these  
12 results suggest that the distance to the core distribution may have a great relevance in  
13 the expansion process (van den Bosch et al., 1992; Acevedo et al., 2005). For instance,  
14 Sullivan et al. (2012) showed that explicitly including dispersal probabilities into the  
15 models yielded better predictions of the future distribution of the species. On the other  
16 hand, the results highlight that models trained with environmental predictors but  
17 without spatial variables can account for the spatial structure of the distributions and the  
18 expansion processes. Lastly, they call into question the usefulness of SDMs - at least in  
19 the context of our study.

20 Range expansion is a highly dynamic process not only determined by the interplay of  
21 demographic and dispersal phenomena, but also by biotic interactions and landscape  
22 complexities (Hastings et al., 2005). Also, the distribution of large-bodied endothermic  
23 and ecologically plastic species (such as the studied species; Acevedo et al., 2005;  
24 Aragón et al., 2010) may prove difficult to predict. All this could explain the absence of  
25 a clear superiority of SDMs over GD to predict range expansion, but still a more

1 profound reason may underlie these results. Spatial autocorrelation allows any spatial  
2 pattern to be modelled using spatially structured variables (Bahn and McGill, 2007).  
3 This is especially true in geographic domains with environmental variations that show a  
4 marked geographic gradient such as Britain (Metzger et al., 2005). The fact that the  
5 inclusion of latitude and longitude as predictors did not significantly change the results  
6 of SDMs proves the latter. This makes it very difficult if not impossible to obtain a  
7 model that does not provide good performance in the training data. In addition, because  
8 of the spatial autocorrelative nature of both the expansion process and the  
9 environmental variables (Legendre, 1993; Diniz-Filho et al., 2003), SDMs behaved  
10 better than chance in T2; the localities that were occupied in T2 had higher suitability  
11 values than the localities that still remained unoccupied. Those localities with higher  
12 suitability values were also the localities that were closer to the range limit in the first  
13 period. Thus, SDMs were not better than GD, which is a much more parsimonious and  
14 preferable model.

15 This study shows that caution should be exercised when using SDMs, and in concrete,  
16 when applying them to forecast range expansion of species that are in disequilibrium  
17 with the environment. By no means do we deny the potential usefulness of SDM or  
18 make the naïve suggestion that GD is a satisfactory way to approximate an expansion  
19 process. It is recognized more and more that, ideally, static correlative SDMs should be  
20 hybridized with dynamic process-based models in order to enhance understanding and  
21 prediction (Brook et al., 2009; Sullivan et al., 2012). Also, SDMs may improve by  
22 accounting for autocorrelation (e.g. Václavík et al., 2012) or by relying on ecological  
23 knowledge rather than on automatic procedures to select the predictors (e.g. Rödder et  
24 al., 2009). However, the point that we wanted to raise in this study was that had not GD  
25 been applied, and given that the SDMs worked better than chance, one could have been

1 tempted to derive misleading conclusions. We have shown that a model that performs  
2 better than chance may lack real significance; by this we mean that either if the interest  
3 is on prediction or in explanation, the most parsimonious model is desirable since the  
4 predictions or the explanation it provides will be more general (Anderson, 2008).  
5 Spatial autocorrelation probably makes the overestimation of models performance be a  
6 rule in SDM (see Hijmans, 2012). If this is so, their significance and usefulness may  
7 also be overvalued. As SDM is currently undergoing exponential growth, we simply  
8 wish to draw attention to its potential misunderstanding and misuse when applied in an  
9 uncritical manner.

10

## 11 **Acknowledgements**

12 We are grateful to two anonymous referees for their useful comments and suggestions  
13 on previous versions of the manuscript. Lucía Maltez kindly reviewed the English. M.  
14 R.-R. was supported by project POIII10-0076-4195 of JCCM, A.J.-V. by the MEC Juan  
15 de la Cierva Program and P. A. was funded from the SFRH/BPD/90320/2012 post-  
16 doctoral grant by Portuguese Fundação para a Ciência e a Tecnologia (FCT) and  
17 European Social Fund.

18

## 19 **References**

20 Acevedo, P., Delibes-Mateos, M., Escudero, M.A., Vicente, J., Marco, J. and Gortázar,  
21 C., 2005. Environmental constraints in the colonization sequence of roe deer  
22 (*Capreolus capreolus* Linnaeus, 1758) across the Iberian Mountains, Spain.  
23 *Journal of Biogeography*, 32: 1671-1680.

24

25

26

27

28

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Acevedo, P., Ward, A.I., Real, R. and Smith, G.C., 2010. Assessing biogeographical  
2 relationships of ecologically related species using favourability functions: a case  
3 study on British deer. *Diversity and Distributions*, 16: 515-528.

4 Acevedo, P., Farfán, M.A., Márquez, A.L., Delibes-Mateos, M., Real, R. and Vargas,  
5 J.M., 2011. Past, present and future of wild ungulates in relation to changes in  
6 land use. *Landscape Ecology*, 26: 19-31.

7 Anderson, D. R., 2008. *Model Based Inference in the Life Sciences. A Primer on*  
8 *Evidence*. Springer, NY.

9 Apollonio, M., Andersen, R. and Putman, R., 2010. *European Ungulates and Their*  
10 *Management in the 21st Century*. Cambridge University Press, Cambridge.

11 Aragón, P., Lobo, J.M., Olalla-Tárraga, M.A. and Rodríguez, M.A., 2010. The  
12 contribution of contemporary climate to ectothermic and endothermic vertebrate  
13 distributions in a glacial refuge. *Global Ecology and Biogeography*, 19: 40-49.

14 Araújo, M.B., Pearson, R.G., Thuiller, W. and Erhard, M., 2005. Validation of species-  
15 climate impact models under climate change. *Global Change Biology*, 11: 1504-  
16 1513.

17 Bahn, V. and McGill, B.J., 2007. Can niche-based distribution models outperform  
18 spatial interpolation? *Global Ecology and Biogeography*, 16: 733-742.

19 Brook, B.W., Akçakaya, H.R., Keith, D.A., Mace, G.M., Pearson, R.G. and Araújo,  
20 M.B., 2009. Integrating bioclimate with population models to improve forecasts  
21 of species extinctions under climate change. *Biology Letters*, 5: 723-725.

22 Cassinello, J., Acevedo, P. and Hortal, J., 2006. Prospects for population expansion of  
23 the exotic aoudad (*Ammotragus lervia*; Bovidae) in the Iberian Peninsula: clues  
24 from habitat suitability modelling. *Diversity and Distributions*, 12: 666-678.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 De Marco, P.Jr., Diniz-Filho, J.A.F and Bini, L.M., 2008. Spatial analysis improves  
2 species distribution modelling during range expansion. *Biology Letters*, 4: 577-  
3 580.  
4  
5  
6  
7 Diniz-Filho, J. A. F., Bini, L. M. and Hawkins, B. A., 2003. Spatial autocorrelation and  
8 red herrings in geographical ecology. *Global Ecology and Biogeography*, 12: 53-  
9 64  
10  
11  
12  
13  
14 European Environment Agency, 2000. Natural resources. CD-ROM. European  
15 Environment Agency, Copenhagen.  
16  
17  
18  
19 Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of  
20 prediction errors in conservation presence/absence models. *Environmental*  
21 *Conservation*, 24: 38-49.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A., 2005. Very high  
2 resolution interpolated climate surfaces for global land areas. *International*  
3 *Journal of Climatology*, 25: 1965-1978.

4 Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N., 2002. Ecological-niche factor  
5 analysis: How to compute habitat-suitability maps without absence data?  
6 *Ecology*, 83: 2027-2036.

7 Hosmer, D.W. and Lemeshow, S., 2000. *Applied Logistic Regression*. Wiley  
8 Interscience, New York.

9 Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating  
10 characteristic curve (AUC) as a discrimination measure in species  
11 distributionmodelling. *Global Ecology and Biogeography*, 21: 498-507.

12 Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P. and Lobo,  
13 J.M., 2011. Use of niche models in invasive species risk assessments. *Biological*  
14 *Invasions*, 13: 2785-2797.

15 Jiménez-Valverde, A., Acevedo, P., Barbosa, A. M., Lobo, J. M. and Real, R., 2013.  
16 Discrimination capacity in species distribution models depends on the  
17 representativeness of the environmental domain. *Global Ecology and*  
18 *Biogeography*, in press.

19 Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:  
20 1659-1673.

21 Lobo, J.M., Jiménez-Valverde, A. and Real, R., 2008. AUC: a misleading measure of  
22 the performance of predictive distribution models. *Global Ecology and*  
23 *Biogeography*, 17: 145-151.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. and Thuiller, W., 2009.  
2 Evaluation of consensus methods in predictive species distribution modelling.  
3 Diversity and Distributions, 15: 59-69.  
4 Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Múcher, C. A. and Watkins, J. W.,  
5 2005. A climatic stratification of the environment of Europe. Global Ecology  
6 and Biogeography, 14: 549-563.  
7 Muñoz, A.R. and Real, R., 2006. Assessing the potential range expansion of the exotic  
8 monk parakeet in Spain. Diversity and Distributions, 12: 656-665.  
9 Peterson, A.T., 2003. Predicting the geography of species invasion via ecological niche  
10 modelling. The Quarterly Review of Biology, 78: 419-433.  
11 Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E.,  
12 Nakamura, M. and Araújo, M.B., 2011. Ecological Niches and Geographic  
13 Distributions. Princeton University Press, Princeton.  
14 Phillips, S.J. and Dudík, M., 2008. Modeling of species distributions with Maxent: new  
15 extensions and a comprehensive evaluation. Ecography, 31: 161-175.  
16 Phillips, S.J., Anderson, R.P. and Schapire, R.E., 2006. Maximum entropy modeling of  
17 species geographic distributions. Ecological Modelling, 190: 231-259.  
18 Rödder, D., Schmidlein, S., Veith, M. and Lötters, S., 2009. Alien invasive slider turtle  
19 in unpredicted habitat: a matter of niche shift or of predictors studied? PLoS  
20 ONE, 4, e7843  
21 Smith, A.B., 2013. On evaluating species distribution models with random background  
22 sites in place of absences when test presences disproportionately sample suitable  
23 habitat. Diversity and Distributions, DOI: 10.1111/ddi.12031



- 1 Sullivan, M.J.P., Davies, R.G., Reino, L. and Franco, A.M.A., 2012. Using dispersal  
2 information to model the species–environment relationship of spreading non-  
3 native species. *Methods in Ecology and Evolution*, 3: 870-879.
- 4 Thuiller, W., Lafourcade, B., Engler, R and Araújo, M.B., 2009. BIOMOD - a platform  
5 for ensemble forecasting of species distributions. *Ecography*, 32: 369-373.
- 6 Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. and Kadmon, R., 2007. A comparative  
7 evaluation of presence-only methods for modelling species distribution.  
8 *Diversity and Distributions*, 13: 397-405.
- 9 Václavík, T., Kupfer, J.A. and Meentemeyer, R.K., 2012. Accounting for multi-scale  
10 spatial autocorrelation improves performance of invasive species distribution  
11 modelling (iSDM). *Journal of Biogeography*, 39: 42-59.
- 12 Vaughan, I.P. and Ormerod, S.J., 2005. The continuing challenges of testing species  
13 distribution models. *Journal of Applied Ecology*, 42: 720-730.
- 14 van den Bosch, F., Hengeveld, R. and Metz, J.A.J., 1992. Analyzing the velocity of  
15 animal range expansion. *Journal of Biogeography*, 19: 135-150.
- 16 Ward, A.I., 2005. Expanding ranges of wild and feral deer in Great Britain. *Mammal  
17 Review*, 35: 165-173.
- 18 Ward, A.I., Etherington, T. and Ewald, J., 2008. Five years of change. *Deer*, 14: 17-20.
- 19 Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, Ş., Amis, M.A., Henderson, L.  
20 and Thuiller, W., 2007. Residence time and potential range: crucial  
21 considerations in modelling plant invasions. *Diversity and Distributions*, 13: 11-  
22 22.
- 23 Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS  
24 Predicting Species Distributions Working Group., 2008. Effects of sample size

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 on the performance of species distribution models. *Diversity and Distributions*,  
2 14: 763-773.

3 Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. and Smith, G.M., 2009. *Mixed*  
4 *effects models and extensions in ecology with R*. Springer, New York.

5

6

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 **Supporting information**

2 **Appendix A** Species distribution modelling: the techniques.

3 **Appendix B** Suitability maps for each studied species and technique in Great Britain.

4 **Appendix C** Receiver Operating Characteristic (ROC) curves for each technique and  
5 species.

6

1 **Table 1** Variables used to model the distribution of the six wild ungulates.

2

Factors	Variables
Climate*	BIO1: Annual mean temperature (°C×10)
	BIO2: Mean diurnal range (°C×10)
	BIO3: Isothermality (BIO2/BIO7)×100 (°C×10)
	BIO4: Temperature seasonality (standard deviation×100)
	BIO5: Max temperature of warmest month (°C×10)
	BIO6: Min temperature of coldest month (°C×10)
	BIO7: Temperature annual range (BIO5– BIO6) (°C×10)
	BIO8: Mean temperature of wettest quarter (°C×10)
	BIO9: Mean temperature of driest quarter (°C×10)
	BIO10: Mean temperature of warmest quarter (°C×10)
	BIO11: Mean temperature of coldest quarter (°C×10)
	BIO12: Annual precipitation (mm)
	BIO13: Precipitation of wettest month (mm)
	BIO14: Precipitation of driest month (mm)
	BIO15: Precipitation seasonality (coefficient of variation)
	BIO16: Precipitation of wettest quarter (mm)
	BIO17: Precipitation of driest quarter (mm)
	BIO18: Precipitation of warmest quarter (mm)
	BIO19: Precipitation of coldest quarter (mm)
Topography <sup>#</sup>	Range of altitude (m)
	Mean altitude (m above sea level)
	Max altitude (m above sea level)
	Mean slope (°)
	Max slope (°)
Spatial	Longitude (m)
	Latitude (m)

3

4 \*Bioclimatic variables were available at ~1 km<sup>2</sup> pixel width from the Worldclim project

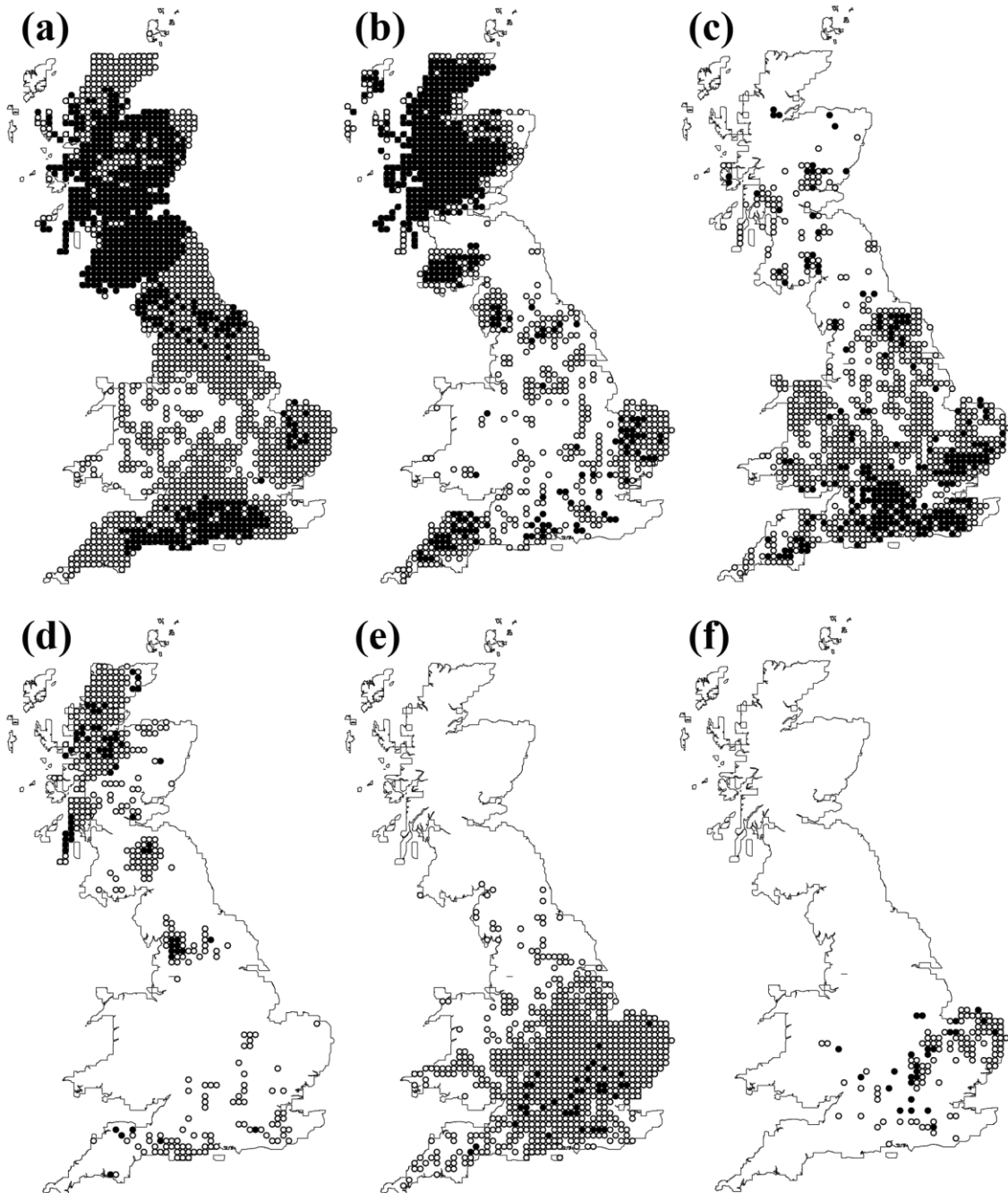
5 database (details in Hijmans et al., 2005). <sup>#</sup>Topographic variables were extracted from

6 the European Digital Elevation Model carried out by the Shuttle Radar Topography

7 Mission (European Environment Agency, 2000) with a spatial resolution of 100 m.

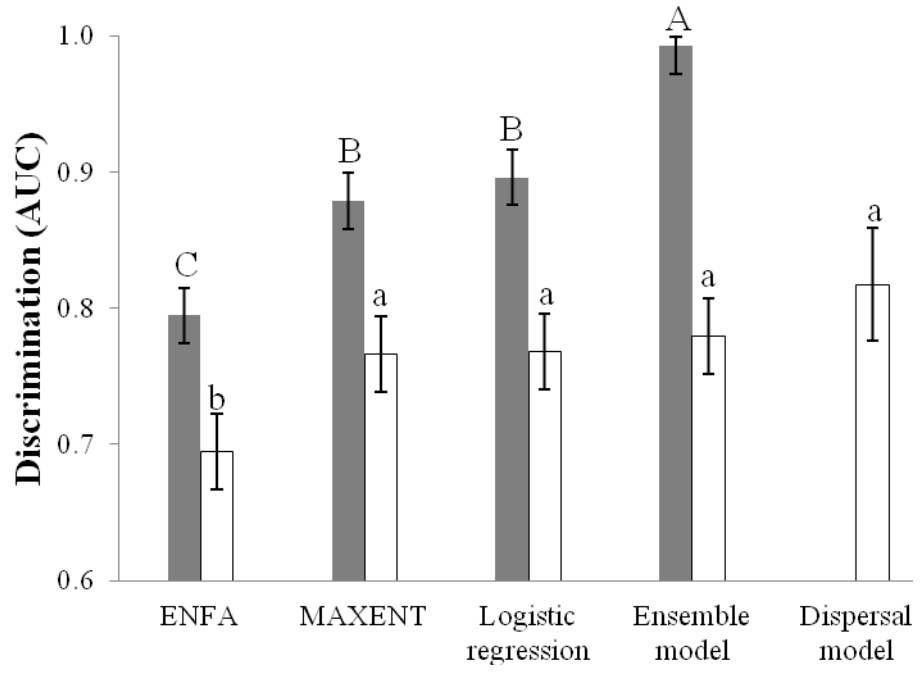
8

1 **Figure 1** Distribution and range expansion of six deer species in Britain between 1972  
2 and 2002. Black circles, observations made up to 1972; open circles, observations made  
3 between 1972 and 2006. (a): roe deer (*Capreolus capreolus*), (b): red deer (*Cervus*  
4 *elaphus*), (c): fallow deer (*Dama dama*), (d): sika deer (*Cervus nippon*), (e): Reeves'  
5 Muntjac (*Muntiacus reevesi*) and (f): Chinese water deer (*Hydropotes inermis*). Adapted  
6 from Ward (2005) and Acevedo et al. (2010).



7  
8

1 **Figure 2** Discrimination capacity (AUC) of the modelling techniques. Techniques were  
 2 assessed i) on the localities occupied up to 1972 and the unoccupied ones (training data,  
 3 T1; in grey), and ii) on the new localities occupied between 1972 and 2006 and the still  
 4 unoccupied ones (T2; in white). Values represent the estimated marginal means (95%  
 5 confidence intervals) obtained from general linear mixed models. Bars sharing the same  
 6 letters (T1, capital; T2, lowercase) indicate techniques that did not significantly differ  
 7 ( $p > 0.05$ ) according to a post-hoc Tukey's test.



9

**Supplementary material for online publication only**

**[Click here to download Supplementary material for online publication only: SUPPORTING INFORMATION\\_Deers\\_UK\\_EcoIMod\\_L](#)**