

1 **TITLE PAGE**

2

3 **Article title:** The devil is in the detail: unstable response functions in species
4 distribution models challenge bulk ensemble modelling

5

6 **Authors:** Henrik Hannemann¹, Katherine J. Willis^{1, 2, 3}, and Marc Macias-Fauria^{4,1}

7

8 ¹ Long-term Ecology Laboratory, Biodiversity Institute, Department of Zoology,
9 University of Oxford, United Kingdom

10 ² Royal Botanic Gardens, Kew, United Kingdom

11 ³ Department of Biology, University of Bergen, Bergen, Norway

12 ⁴ School of Geography and the Environment, University of Oxford, United Kingdom

13 **Research address:** Long-term Ecology Laboratory, Biodiversity Institute,
14 Department of Zoology, University of Oxford, Tinbergen Building, South Parks
15 Road, Oxford, OX1 3PS, United Kingdom.

16 **Email addresses of authors:** henrik.hannemann@zoo.ox.ac.uk

17 kathy.willis@zoo.ox.ac.uk

18 marc.maciasfauria@ouce.ox.ac.uk

19 **Running title:** Unstable response functions in SDMs

20 **Correspondence author:** Henrik Hannemann

21 **Word count (Abstract):** 300

22 **Word count (Main body and biosketch):** 5,198 (incl. references and biosketches)

23 **Number of references:** 45

24

ABSTRACT

Aim: Species distribution models (SDMs) are commonly used to determine biodiversity threats and opportunities under climate change. Despite SDMs being based on the assumption of complete knowledge of the climate space of the modelled species, truncated occurrence datasets (and hence truncated climate spaces) such as national inventories are often employed. This may lead to prediction errors, which have been proposed to stem from 1) the degree of climate space truncation and/or 2) instability of modelling algorithms. Our aim was to explore potential causes of prediction errors in SDMs based on using truncated training data sets.

Location: Europe 11°W–32°E, 34°–72°N

Methods: SDMs were applied to seven forest tree species employing commonly used bioclimatic variables. We created two model training data sets covering 1) Germany only (significantly truncated climate space) and 2) Europe (minimally truncated climate space). Differences between the climate space represented by Germany-only and European data were measured on two-dimensional climate spaces obtained through principal component analysis of the bioclimatic variables. Seven SDM algorithms were run, and response function stability and variable selection for each species and model type were analysed.

Results: The degree of climate space truncation was less important for model performance than the instability of model algorithms and indiscriminate variable

selection. The latter led to irrelevant relationships of species occurrence with bioclimatic variables. These instabilities caused pronounced prediction errors.

Main conclusions: Our results strongly suggest that erroneous model predictions stem from instability and ecological irrelevance of the statistical functions relating species probability of occurrence with bioclimatic variables, compounded with a lack of consistency in variable selection. Models displaying these characteristics showed lower overall performance when trained with truncated data sets. Further, commonly used ensemble approaches do not compensate the shortfalls of individual models. Detailed model-by-model, species-by-species analysis of response functions and variable importance is recommended.

Keywords: climate space, conservation planning, Europe, forest tree species, response function, species distribution modelling, truncated niche space, ensemble modelling

INTRODUCTION

In order to successfully conserve biodiversity, the potential impact of climatic change on species needs to be accounted for, and both current and future distributions assessed (Araújo & Rahbek, 2006). Species distribution modelling (SDM, a.k.a. environmental niche modelling or bioclimatic modelling) is frequently employed to determine changes in species' distribution patterns in response to climate change and to assess potential threats – or resilience – in the light of loss of suitable climate space (Thomas *et al.*, 2004; Araújo *et al.*, 2011; Milad *et al.*, 2011). Such models are routinely used to provide insight into large-scale ecological patterns and niche behaviour under environmental or climatic change (Thuiller *et al.*, 2009).

Due to their comparatively easy implementation and low data requirements, SDMs are often utilised in 'bulk' modelling of vast numbers of species (Araújo *et al.*, 2011; Thuiller *et al.*, 2014). They are based on the notion of environmental niche as conceptualised by Hutchinson (1957) and are built by statistically linking species occurrence and environmental (most often bioclimatic) data, identifying an n -dimensional space contained within a set of relevant and preferential climate conditions commonly termed as the 'climate space' of the species. Once fitted (or calibrated, or trained), models can be projected onto other datasets comprising environmental/bioclimatic parameters for future or spatially different scenarios in order to study potential spatial and/or temporal distributional changes (e.g. in case of future distribution projections or invasive species acquiring new ranges; Carvalho *et al.*, 2011; Pearman *et al.*, 2008; Spangenberg *et al.*, 2012; Thuiller *et al.*, 2008).

89 A basic assumption in setting up SDMs is that the full climate space of the
90 modelled species is observed and incorporated into the training dataset. Yet, in many
91 cases SDMs are constructed using data derived from reduced datasets (e.g. individual
92 countries) that may not coincide with their full present geographical ranges, and/or
93 occurrence data may be incomplete. As a consequence, conservation
94 recommendations stemming from their use are often based on a truncated bioclimatic
95 niche space, which is used to infer current and future distribution areas regardless
96 (Broadmeadow, 2005; Bálint *et al.*, 2011; Edman *et al.*, 2011; Keenan *et al.*, 2011;
97 Bertrand *et al.*, 2012; Raes, 2012; Serra-Diaz *et al.*, 2012). The extent to which the
98 use of these truncated datasets may influence conservation recommendations is, as
99 yet, unknown.

101 Palaecological studies on past distributions of several taxa highlight the very
102 distinct possibility that even in the case where the full present geographical range of a
103 given species is known, its full bioclimatic space might remain unknown (e.g. Veloz
104 *et al.*, 2012). This might be due to anthropogenic activities (Willis & Birks, 2006),
105 insufficient post-refugial dispersal time, physical dispersal barriers (Soberón, 2007),
106 or to the ability of the species to tolerate climatic conditions that are not observed now
107 but that occurred in the past (non-analogous climate; Williams and Jackson, 2007).
108 Thus, even in best-case scenarios, we are left with uncertainty as to this basic
109 assumption in SDMs, making it a more general problem than is often realised.

111 Poor model performance stemming from the use of training data sets where
112 the full bioclimatic range of the species is not included is thought to result from 1) the
113 inability of statistical SDMs' to forecast species occurrence probabilities into non-

analogous climate space (e.g. Harrison *et al.*, 2006; Rodríguez-Castañeda *et al.*, 2012; Williams and Jackson, 2007) and/or 2) the instability of model algorithms, as seen in unstable response functions – where the true relationship between the probability of species occurrence and climate is not properly captured – and in arbitrary variable selection (Thuiller, 2004; Thuiller *et al.*, 2004; Araújo *et al.*, 2005b; Barbet-Massin *et al.*, 2010). In the former, maximizing as much as possible the climate space used in the training data would reduce the chance of SDM errors, whereas in the latter the reasons for response function instability and arbitrary variable selection – and whether they are at all related to truncation of the species climate space in the training data – would need to be established. Thus, although being widely discussed (e.g. Soberón, 2007), the statistical causes of such errors have not been extensively studied, especially when using (as often is the case) truncated datasets (e.g. Keenan *et al.*, 2011; Serra-Diaz *et al.*, 2012; Thuiller *et al.*, 2014).

Ensemble modelling has been proposed as a compromise in order to encompass some of these issues (Araújo & New, 2007). This approach consists of combining output from multiple models – by e.g. averaging predicted occurrence probabilities – to provide more robust projections or forecasts of distributions, and stems from the ensemble methods employed in climate modelling (Araújo & New, 2007; Thuiller *et al.*, 2009). The ensemble modelling approach assumes that individual model errors are minimised in favour of a common inter-model signal, and that averaging potentially very different response functions will result in clearer model predictions. However, this is based on the assumption that the response functions in each individual model are broadly correct and reflect the relationship between species and bioclimatic variables. If this were not the case, ensemble

averaging of models with unstable response functions could potentially result in an addition of noise.

The aim of this study was thus to understand the shortfalls of species distribution modelling when using truncated training datasets. We evaluated two explanations for prediction errors in SDMs, namely (1) the effect of climate space – i.e. how well do the models perform outside of the bioclimatic niche space represented in a truncated dataset on which they were trained and (2) the effect of model algorithm choice – affecting variable selection and response function stability i.e. how does the use of a truncated dataset influence the description of the relationship between the individual species probability of occurrence and its bioclimatic variables. We then discuss the implications of our results for use of ensemble modelling approaches.

DATA

Continental Europe (EUR, 11°W – 32°E, 34° – 72°N) provides an ideal study area as it contains eight biogeographic regions (Alpine, Atlantic, Black Sea, Boreal, Continental, Mediterranean, Pannonian, and Steppic regions) and thus a strong gradient of bioclimatic factors and a concurrent spatial turnover in species distributions. This diversity of bioclimatic factors is an important requirement for SDMs if they are to be a useful tool for e.g. future-oriented conservation planning in the face of climate change (e.g. Margules & Pressey, 2000; Pearson & Dawson, 2005; Wilson *et al.*, 2005; Moilanen *et al.*, 2009; Carvalho *et al.*, 2011; McMahon *et al.*, 2011). Germany was chosen as a political entity to truncate occurrence data (GER). This is because Germany contains coverage from the Atlantic biogeographic region in

the northwest, the Continental region mainly throughout, and the Alpine region at its southern end, and it therefore covers several bioclimatic zones.

Occurrence data from seven forest tree species with a distribution within Europe and including presence in Germany was obtained from the 'European Forest Genetic Resources Programme' (EUFORGEN, <http://www.euforgen.org/>). These were: silver fir (*Abies alba* Mill.), sycamore (*Acer pseudoplatanus* L.), beech (*Fagus sylvatica* L.), European larch (*Larix decidua* Mill.), pedunculate oak (*Quercus robur* L.), service tree (*Sorbus domestica* L.), and European white elm (*Ulmus laevis* Pall.). If covering a larger area than Europe (as in the case of *A. pseudoplatanus* that has a disjointed population in the Caucasus), *Q. robur* (fragmented population in Western Russia and the Caucasus), and *U. laevis* (continuous distribution in Western Russia), the distribution was minimally clipped according to the designated extent of Europe. EUFORGEN supplies presence and absence data as high-resolution shapefiles which were gridded according to the 10' spatial grid used by the bioclimatic variables. This created a presence/absence matrix throughout Europe for each individual species (Table 1; 30,430 presence and absence data points).

Bioclimatic data for 1901 – 2000 was obtained from the Climate Research Unit (CRU, <http://www.cru.uea.ac.uk/>). Widely used bioclimatic variables reflecting ecological requirements for forest tree species were employed in the modelling exercise (Prentice *et al.*, 1992). These were: annual precipitation (mm), summer precipitation (mm; June, July, August), winter precipitation (mm; December, January, February), mean annual temperature (°C), minimum temperature (°C), mean summer

temperature (°C; June, July, August), growing degree days (GDD; days above 5°C), and equilibrium evapotranspiration (mm).

METHODS

Seven statistical algorithms commonly used in SDM were employed in this study. These were: artificial neural networks (ANN), classification tree analysis (CTA), generalised additive model (GAM), generalised boosted model (GBM), generalised linear model (GLM), multiple adaptive regression spline (MARS), and random forest (RF). In these models, each species is represented by its n -dimensional climate space delineated by the bioclimatic variables. Models were run separately using non-truncated (EUR) and truncated (GER) training data sets. The BIOMOD R package (Thuiller *et al.*, 2009) was used to run the model algorithms.

In order to avoid collinearity between the model covariates, only growing degree days, annual precipitation, and summer precipitation were used (Prentice *et al.*, 1992; Elith & Leathwick, 2009) after a correlation matrix of all the bioclimatic variables was used to exclude highly correlated pairs of bioclimatic variables ($r^2 > 0.65$). Each model run was repeated three times for each species: all models were run using a 70% random sample as calibration and a 30% random sample as validation data (cross-validation) (Breiman, 1993). These random splits differed for each of the iterative runs. These iterative runs are common practice and are designed to provide the best possible insights given the limited data available. Predicted probabilities of occurrence were transformed into presence-absence predictions by selecting a cut-off probability value for each model type and species that would maximise model performance. Area Under the Curve of the Receiver Operating Characteristic statistic

(AUC), Cohen's Kappa, and True Skill Statistic (TSS) were evaluated (Hanley & McNeil, 1982; Monserud & Leemans, 1992; Allouche *et al.*, 2006).

1) The effect of climate space

We hypothesised that models for species whose climate space had been severely truncated by the choice of Germany as a training data set (that is, species with a total climate space encompassing conditions far from those observed in Germany only) would perform worse than models for species whose climate space was more similar to the conditions experience in Germany only. For example, the climate space of a species with a range restricted to central Europe will suffer from a smaller loss of information if only the climatic conditions observed in Germany are considered than that of a species with a range encompassing central and southern Europe, and this loss of information should result in poorer model performance. The association between the degree of truncation in the training data set and model performance was thus tested by comparing the performance of SDMs trained with truncated (GER) and non-truncated (EUR) data sets with the degree of climate space truncation in the training data.

A principal component analysis was computed on all the bioclimatic variables (EUR extent) to reduce the dimensions of the bioclimatic dataset and simplify the calculation of climate space extent and truncation. The two largest principal components were temperature and water availability – accounting respectively for 62.41% and 27.95% of the total variance (i.e. 90.36% of the total variance combined). These two variables were then used to compute climate spaces for both the EUR and the GER datasets (Barbet-Massin *et al.*, 2010) (*Supplementary Materials*, p. 2 and 3).

238

239 For each species, occurrences were projected onto the two-dimensional
240 climate space. We applied the following measurements: percentage overlap between
241 EUR and GER climate space areas (defined as the convex hull of the climate space),
242 percentage overlap between EUR and GER occurrence points, distance between the
243 EUR and GER climate space centres of density, and distance between the EUR and
244 GER geometric centroids. Measurements were then repeated with rarefied datasets in
245 order to exclude potential bias in natural climate space availability of the occurrence
246 data (Broennimann *et al.*, 2012).

247

248 **2) The effect of model algorithm choice**

249 The choice of the modelling algorithm has been shown to affect model performance
250 and predicted distributions (Araújo *et al.*, 2005a). Model algorithms determine the
251 characteristics of the response functions, which attribute values of probability of
252 species presence within the climate space. Likewise, model algorithms determine the
253 selection of and the weight given to the bioclimatic variables used as predictors for
254 species occurrence. Poorly performing models may show erratic response curves and
255 variable selection. We thus assessed response curves and variable selection for each
256 individual model and species for three scenarios: GER projected onto GER
257 (significantly truncated, analogous); GER onto EUR (significantly truncated, non-
258 analogous); and EUR on EUR (minimally truncated, analogous).

259

260 Response functions were examined for each individual model and for each
261 species and scenario respectively as well as averaged for the species on the whole in
262 an ensemble approach (Araújo & New, 2007). They consist of functions that estimate,

for each species and bioclimatic variable, probabilities of occurrence along the climatic gradients defined by the range of climatic conditions observed in the training data sets. The ensemble projection consisted of a weighted average of model outputs according to model performance statistics. In order to maximise sensitivity-specificity sums and minimise their difference, the ensemble models were weighted according to AUC (Hanley and McNeil, 1982).

Variable selection and weighting can be assessed by the statistic ‘variable importance’, which is calculated as “1- (*the correlation between the estimated probability of presence made by a model including all variables except for a randomised target variable, and a model employing all variables*)” (Thuiller *et al.*, 2009). According to this, low variable importance scores (e.g. < 0.3) indicate high degrees of correlation between the two model predictions – that is, the inclusion or exclusion of the target variable does not affect the overall model forecast, indicating low variable importance, and vice versa. Variables showing low importance typically have very low weight/loadings and hence do not contribute to SDM predictive ability.

The joint analysis of response functions and variable importance provides information on the effects of the choice of model algorithm. In the worst case scenario, a model algorithm might produce response functions not reflecting any real relationship between a bioclimatic variable and the probability of occurrence of a species (e.g. overfitting), and at the same time attribute a large importance to this bioclimatic variable, magnifying the biased response function and resulting into completely erroneous predicted species distributions.

RESULTS

1) The effect of climate space

Model performance statistics did not show any consistent relationship with the proportion of climate space occupied by the truncated GER vs. the non-truncated EUR data sets, as measured by distances between centres of gravity, geometric centroids, or area or presence overlap (*Table 2*). This suggests that the degree of climatic truncation in the model training data was not a decisive factor to explain prediction uncertainties.

2) The effect of model algorithm choice

Response function instability and variable importance were shown to contribute to important differences in model performance for all seven species used and thus to influence final predictions of species occurrence, especially when using a truncated data set.

Examples highlighting these findings are illustrated in the output for the following species:

- i) *L. decidua* displayed consistent results with largely accurate predictions both under truncated (GER) and full-scale (analogous – EUR-trained projected on EUR – and non-analogous – GER-trained projected on EUR) bioclimatic training data (*Figure 1, Table 2*). Each of the individual models suggested robust variable selection identifying summer precipitation as key in both truncated and non-truncated projections (*Figure 2A*). Although the spread of predictions increased in wetter summer conditions, all models agreed in assigning a zero probability of *L. decidua* presence in dry summer locations,

and suggested a threshold along the summer precipitation gradient (*Figures 2A, C and Supplementary Materials p. 15*). Although the response functions of the remaining model bioclimatic variables portrayed erratic behaviour, variable importance for these variables has very low (e.g. 0.013 for GDD, 0.034 for annual precipitation, vs. 1.014 for summer precipitation in GER training data, *Table 1 and Figure 2A*). This implied that the importance of summer precipitation and its stable response functions was able to override the influence of erratic response functions for unimportant variables in the final model prediction.

- ii) The distribution of *A. pseudoplatanus* was largely successfully predicted in its current occurrence range throughout Europe when models were trained over EUR (analogous). However, large-scale false-positive forecasts were observed when a non-analogous projection was used (i.e. GER-trained projected onto Europe; *Figure 3A, Table 2*). Both individual and ensemble response functions showed erratic behaviour (best seen in *Figure 2B, D, and Supplementary Materials, p. 5-11*). Variable importance was also inconsistent between each of the separate model runs and types. No variable was identified as particularly important (*Table 1*). The standard error of the ensemble functions remained large, and the overall importance of variables with response functions that might contribute to good predictive ability was too low to mask the noisy predictive ability of the other bioclimatic variables (*Figure 2B*).

Importantly, the response functions for the individual algorithms demonstrated varying patterns even for the same species and the same training data. For example, when comparing models trained on GER and projected on EUR with those both

trained and projected on GER, no stability of the response function was observed, even in the climate range area of GER (*Figure 2* and *Supplementary Material*, p. 5-11).

DISCUSSION

Species distribution modelling is based on the combined understanding of species occurrence data and its relationship to ecologically relevant climatic/environmental conditions in which it exists (its bioclimatic envelope). Whereas the truncation of the climate space – potentially restricting its true bioclimatic envelope – resulted in worse model performance into non-analogue conditions, a surprising result from this study was that poor model performance did not depend on the degree of truncation (i.e. we did not observe a positive or meaningful relationship between the degree of truncation and model performance) (*Table 2*). Instead, it was the combined effect of the response functions (i.e. the shape of the predicted relationship between the probability of species occurrence and its bioclimatic variables) and variable selection by the model algorithms that showed different behaviour in individual model runs, and this was the main factor responsible for poor SDM performance. Within this modelling exercise, the instability of response functions was therefore shown to contribute to significant errors in SDMs and to influence final predictions when using truncated data sets. Thus, the usefulness of SDMs dramatically dropped when the shapes of the functions describing the probability of presence along bioclimatic gradients changed depending on the training datasets used (*Figure 2*). Models showing these characteristics showed the worst predictive performance when trained with truncated datasets.

The characteristics of individual model algorithms can be clearly seen when the response functions are analysed (e.g. whereas algorithms based on smoothing functions will show smooth response functions, decision tree-based algorithms will tend to present step-like response functions). This not only allows the identification of common statistical fit, but also the assessment of ecologically sound threshold changes or optimums (*Figure 2*). Ecological *soundness* might ideally be assessed on the light of prior knowledge of the species ecology (e.g. Oksanen & Minchin, 2002), however identifying unsound response functions does not necessarily require this: an ecologically sound function depicting the probability of occurrence for a given species across a bioclimatic gradient should not consist of several optima and minima, nor show strongly irregular shapes. This should be a warning for overfitting, or at least for model fitting of underlying processes beyond the explanatory reach of the bioclimatic variable being used (Oksanen & Minchin, 2002; Austin, 2007; Townsend Peterson, 2011). Overall, the shape and structure of the response functions allow an initial assessment of ecologically sound statistical model fit.

Although no clear cause for the erratic behaviour of the response functions was identified, response function stability (*Figure 2A, B*) coincided with variable selection; i.e. variables identified as important had more consistent response functions throughout all training data regimes and model types as shown by e.g. *L. decidua*. This would therefore suggest that a process to identify important variables can help to suppress erratic variable responses and allow successful prediction as in the case of *L. decidua*, for which summer precipitation (*Figure 2A*) has by far the highest influence and thus the irregular response functions of other variables are proportionally of no importance.

We also found that aggregation into ensemble predictions (see *Figure 2A, B, Supplementary Materials*, p. 5-11; Araújo & New, 2007; Thuiller *et al.*, 2009) did not necessarily results in improved predictive ability. We examined all three available statistics (AUC, Cohen's Kappa, TSS; *Table 2*) as threshold values to optimise model performance. Regardless of the use of truncated or non-truncated occurrence data, all ensemble models indicated shortfalls. In many cases, there was no clear constraining or promoting signal for any of the species; instead an overall noisy prediction resulted due to the averaging of a series of data fitting exercises. Further noise was added to the ensemble models prediction through inclusion of models with indiscriminate variable importance. Ensemble averaging does not therefore necessarily provide more reliable predictions, and especially under non-analogous conditions, meaningless values of variable importance may lead to erroneous predictions. In this respect, the main limitation in the analogy between ensemble projections of physically-based climate models and those of statistical SDMs relies on the *blindness* of SDMs to capture ecological processes linking climate and species occurrence: in order to determine whether a given SDM might be capturing an ecological process and thus can be incorporated in an ensemble averaging, a close and critical inspection of its characteristics is needed.

Our results suggest that an optimal approach to robust SDM use lies in the assessment of the across-model consistency and a good understanding of the response functions, paired with the appropriate weighting of variables used within the modelling process. Given that SDM modelling is still the choice method used among other things to assess the threat potential to biodiversity (Thomas *et al.*, 2004; Araújo

et al., 2011), it is essential that the relationship between the bioclimatic variables and the probability of species occurrence is captured in consistent response functions, and that these factors are taken into consideration, especially before truncated datasets or projections into non-analogue climate spaces are employed.

Our results strongly support the idea of conservation planning schemes based on robust model results paired with the ecological necessity to move away from restricted planning scopes (e.g. political boundaries as represented by the GER data set). We have shown that prediction errors stem from the individual algorithms used and are readily identifiable within single predictions. Overall we suggest that more stringent principles need to be applied in order to supply more reliable and ecologically realistic modelling results. The large-scale prediction errors shown in this study highlight the importance of an integrated research approach in order to connect conservation tools with ecological realities.

ACKNOWLEDGEMENTS

The Oxford Martin Programme on Resource Stewardship (University of Oxford) funded the work of MM-F.

430 REFERENCES

- 431 Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species
432 distribution models: prevalence, kappa and the true skill statistic (TSS).
433 *Journal of Applied Ecology*, **43**, 1223–1232.
- 434 Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D. & Thuiller, W. (2011)
435 Climate change threatens European conservation areas. *Ecology Letters*, **14**,
436 484–492.
- 437 Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions.
438 *Trends in Ecology & Evolution*, **22**, 42–47.
- 439 Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005a) Validation of
440 species-climate impact models under climate change. *Global Change Biology*,
441 **11**, 1504–1513.
- 442 Araújo, M.B. & Rahbek, C. (2006) How Does Climate Change Affect Biodiversity?
443 *Science*, **313**, 1396–1397.
- 444 Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005b) Downscaling
445 European species atlas distributions to a finer resolution: implications for
446 conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.
- 447 Austin, M. (2007) Species distribution models and ecological theory: A critical
448 assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–
449 19.
- 450 Bálint, M., Domisch, S., Engelhardt, C.H.M., Haase, P., Lehrian, S., Sauer, J.,
451 Theissinger, K., Pauls, S.U. & Nowak, C. (2011) Cryptic biodiversity loss
452 linked to global climate change. *Nature Climate Change*, **1**, 313–318.
- 453 Barbet-Massin, M., Thuiller, W. & Jiguet, F. (2010) How much do we overestimate
454 future local extinction rates when restricting the range of occurrence data in
455 climate suitability models? *Ecography*, **33**, 878–886.
- 456 Bertrand, R., Perez, V. & Gégout, J.-C. (2012) Disregarding the edaphic dimension in
457 species distribution models leads to the omission of crucial spatial information
458 under climate change: the case of *Quercus pubescens* in France. *Global
459 Change Biology*, **18**, 2648–2660.
- 460 Breiman, L. (1993) *Classification and regression trees*, Chapman & Hall, New York.
- 461 Broadmeadow, M.S.J. (2005) Climate change and the future for broadleaved tree
462 species in Britain. *Forestry*, **78**, 145–161.
- 463 Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L.,
464 Yoccoz, N.G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N.E.,
465 Graham, C.H. & Guisan, A. (2012) Measuring ecological niche overlap from
466 occurrence and spatial environmental data. *Global Ecology and Biogeography*,
467 **21**, 481–497.
- 468 Carvalho, S.B., Brito, J.C., Crespo, E.G., Watts, M.E. & Possingham, H.P. (2011)
469 Conservation planning under climate change: Toward accounting for
470 uncertainty in predicted species distributions to increase confidence in
471 conservation investments in space and time. *Biological Conservation*, **144**,
472 2020–2030.
- 473 Edman, T., Angelstam, P., Mikusiński, G., Roberge, J.-M. & Sikora, A. (2011)
474 Spatial planning for biodiversity conservation: Assessment of forest
475 landscapes' conservation value using umbrella species requirements in Poland.
476 *Landscape and Urban Planning*, **102**, 16–23.

- 477 Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological
478 Explanation and Prediction Across Space and Time. *Annual Review of*
479 *Ecology, Evolution, and Systematics*, **40**, 677–697.
- 480 Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver
481 operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
- 482 Harrison, P.A., Berry, P.M., Butt, N. & New, M. (2006) Modelling climate change
483 impacts on species' distributions at the European scale: implications for
484 conservation policy. *Environmental Science & Policy*, **9**, 116–128.
- 485 Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harbor Symposia on*
486 *Quantitative Biology*, **22**, 415–427.
- 487 Keenan, T., Maria Serra, J., Lloret, F., Ninyerola, M. & Sabate, S. (2011) Predicting
488 the future of forests in the Mediterranean under climate change, with niche-
489 and process-based models: CO2 matters! *Global Change Biology*, **17**, 565–
490 579.
- 491 Margules, C.R. & Pressey, R.L. (2000) Systematic conservation planning. *Nature*,
492 **405**, 243–253.
- 493 McMahon, S.M., Harrison, S.P., Armbruster, W.S., Bartlein, P.J., Beale, C.M.,
494 Edwards, M.E., Kattge, J., Midgley, G., Morin, X. & Prentice, I.C. (2011)
495 Improving assessment and modelling of climate change impacts on global
496 terrestrial biodiversity. *Trends in Ecology & Evolution*, **26**, 249–259.
- 497 Milad, M., Schaich, H., Bürgi, M. & Konold, W. (2011) Climate change and nature
498 conservation in Central European forests: A review of consequences, concepts
499 and challenges. *Forest Ecology and Management*, **261**, 829–843.
- 500 Moilanen, A., Wilson, K.A. & Possingham, H.P. eds. (2009) *Spatial conservation*
501 *prioritization: quantitative methods and computational tools*, Oxford
502 University Press, Oxford ; New York.
- 503 Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the
504 Kappa statistic. *Ecological Modelling*, **62**, 275–293.
- 505 Oksanen, J. & Minchin, P.R. (2002) Continuum theory revisited: what shape are
506 species responses along ecological gradients? *Ecological Modelling*, **157**,
507 119–129.
- 508 Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics
509 in space and time. *Trends in Ecology & Evolution*, **23**, 149–158.
- 510 Pearson, R.G. & Dawson, T.P. (2005) Long-distance plant dispersal and habitat
511 fragmentation: identifying conservation targets for spatial landscape planning
512 under climate change. *Biological Conservation*, **123**, 389–401.
- 513 Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon,
514 A.M. (1992) Special Paper: A Global Biome Model Based on Plant
515 Physiology and Dominance, Soil Properties and Climate. *Journal of*
516 *Biogeography*, **19**, pp. 117–134.
- 517 Raes, N. (2012) Partial versus Full Species Distribution Models. *Natureza &*
518 *Conservação*, **10**, 127–138.
- 519 Rodríguez-Castañeda, G., Hof, A.R., Jansson, R. & Harding, L.E. (2012) Predicting
520 the Fate of Biodiversity Using Species' Distribution Models: Enhancing
521 Model Comparability and Repeatability. *PLoS ONE*, **7**, e44402.
- 522 Serra-Diaz, J.M., Ninyerola, M. & Lloret, F. (2012) Coexistence of *Abies alba* (Mill.)
523 – *Fagus sylvatica* (L.) and climate change impact in the Iberian Peninsula: A
524 climatic-niche perspective approach. *Flora - Morphology, Distribution,*
525 *Functional Ecology of Plants*, **207**, 10–18.

- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
- Spangenberg, J.H., Bondeau, A., Carter, T.R., Fronzek, S., Jaeger, J., Jylhä, K., Kühn, I., Omann, I., Paul, A., Reginster, I., Rounsevell, M., Schweiger, O., Stocker, A., Sykes, M.T. & Settele, J. (2012) Scenarios for investigating risks to biodiversity: Investigating risks to biodiversity. *Global Ecology and Biogeography*, **21**, 5–18.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020–2027.
- Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.
- Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Thuiller, W., Pironon, S., Psomas, A., Barbet-Massin, M., Jiguet, F., Lavergne, S., Pearman, P.B., Renaud, J., Zupan, L. & Zimmermann, N.E. (2014) The European functional tree of bird life in the face of global change. *Nature Communications*, **5**.
- Townsend Peterson, A. ed. (2011) *Ecological niches and geographic distributions*, Princeton University Press, Princeton, N.J.
- Veloz, S.D., Williams, J.W., Blois, J.L., He, F., Otto-Bliesner, B. & Liu, Z. (2012) No-analog climates and shifting realized niches during the late quaternary: implications for 21st-century predictions by species distribution models. *Global Change Biology*, **18**, 1698–1713.
- Williams, J.W. & Jackson, S.T. (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475–482.
- Willis, K.J. & Birks, H.J.B. (2006) What Is Natural? The Need for a Long-Term Perspective in Biodiversity Conservation. *Science*, **314**, 1261–1265.
- Wilson, K.A., Westphal, M.I., Possingham, H.P. & Elith, J. (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation*, **122**, 99–112.

570 **BIOSKETCHES**

571 **Henrik Hannemann** is a DPhil student in the Long-term Ecology Laboratory based
572 in the Department of Zoology, University of Oxford. His research interest focuses on
573 conservation planning and effective resource use under environmental change.

574 **Katherine J. Willis** holds a Professorship of Biodiversity in the Department of
575 Zoology, is the Associate Director of the Biodiversity Institute, and a Professorial
576 Fellow at Merton College at the University of Oxford. In addition, Professor Willis is
577 Director of Science at the Royal Botanic Gardens, Kew, United Kingdom, and an
578 adjunct professor (Professor II) in the Department of Biology, University of Bergen,
579 Norway.

580 **Marc Macias-Fauria** is a NERC Independent Research Fellow and Associate
581 Professor in Physical Geography at the School of Geography and the Environment,
582 and a tutorial fellow at St Peter's College at the University of Oxford. His research is
583 directed at understanding the interactions between physical and biological systems
584 over spatial scales ranging from study sites to continents and temporal scales ranging
585 from decades to millennia.

586

	Dataset		EUR training data			GER training data		
	presences	absences	GDD	pre	presu	GDD	pre	presu
<i>A. alba</i>	2,215	28,215	0.269	0.292	0.937	0.171	0.343	1.017
<i>A. pseudoplatanus</i>	7,967	22,463	0.637	0.328	0.635	0.311	0.410	0.614
<i>F. sylvatica</i>	5,871	24,559	0.577	0.440	0.350	0.316	0.899	0.195
<i>L. decidua</i>	1,237	29,193	0.191	0.322	0.985	0.013	0.034	1.014
<i>Q. robur</i>	13,871	16,559	0.681	0.152	0.448	0.984	0.367	0.437
<i>S. domestica</i>	5,579	24,851	0.630	0.392	0.371	0.491	0.645	0.608
<i>U. laevis</i>	10,753	19,677	0.621	0.228	0.658	0.578	0.406	0.769

Table 1 Presence and absence data within the dataset for each species is shown in the first two columns. The following columns depict the importance of each individual variable (growing degree days *GDD*, annual precipitation *pre*, and summer precipitation *presu*) within the ensemble species distribution models for each of the seven species modelled. The importance of each variable was calculated as $1 - (\text{the correlation between the predicted probability of presence made by 1) a model including all variables except for a randomised target variable, and 2) a model employing all variables})$. Note that for some species (e.g. *A. pseudoplatanus*, *F. sylvatica*, *S. domestica*) no clear indication of consistent variable selection can be identified, whereas for some others (e.g. *L. decidua*, *A. alba*) variable importance is consistent across the training data sets.

	AUC	Kappa	TSS	Distance in climate space		Overlap (%)	
				Centre of gravity	Geometric	Area	Points
<i>L. decidua</i>	0.884	0.518	0.704	0.585	1.226	15.608	16.087
<i>S. domestica</i>	0.715	0.257	0.398	1.262	2.212	7.093	5.180
<i>A. alba</i>	0.659	0.117	0.332	0.308	1.751	10.598	19.233
<i>Q. robur</i>	0.641	0.301	0.326	0.253	3.498	6.994	11.391
<i>F. sylvatica</i>	0.629	0.158	0.266	0.097	1.840	12.874	21.240
<i>A. pseudoplatanus</i>	0.597	0.194	0.230	0.242	1.782	11.392	14.924
<i>U. laevis</i>	0.554	0.226	0.257	0.429	1.332	11.169	9.328

599 **Table 2** Measurements of model performance and climate space for all species. AUC,
 600 Cohen's Kappa, and TSS are given for comparison of the ensemble models for GER
 601 training data (Germany only) projected onto Europe (see Data and Methods).
 602 Geometric distance between the centre points of the convex hulls of each species'
 603 training data sets has been calculated. The same convex hull delineation was used to
 604 calculate the climate space area. Note that there is no clear relationship between
 605 measurements of climate space and model performance.
 606

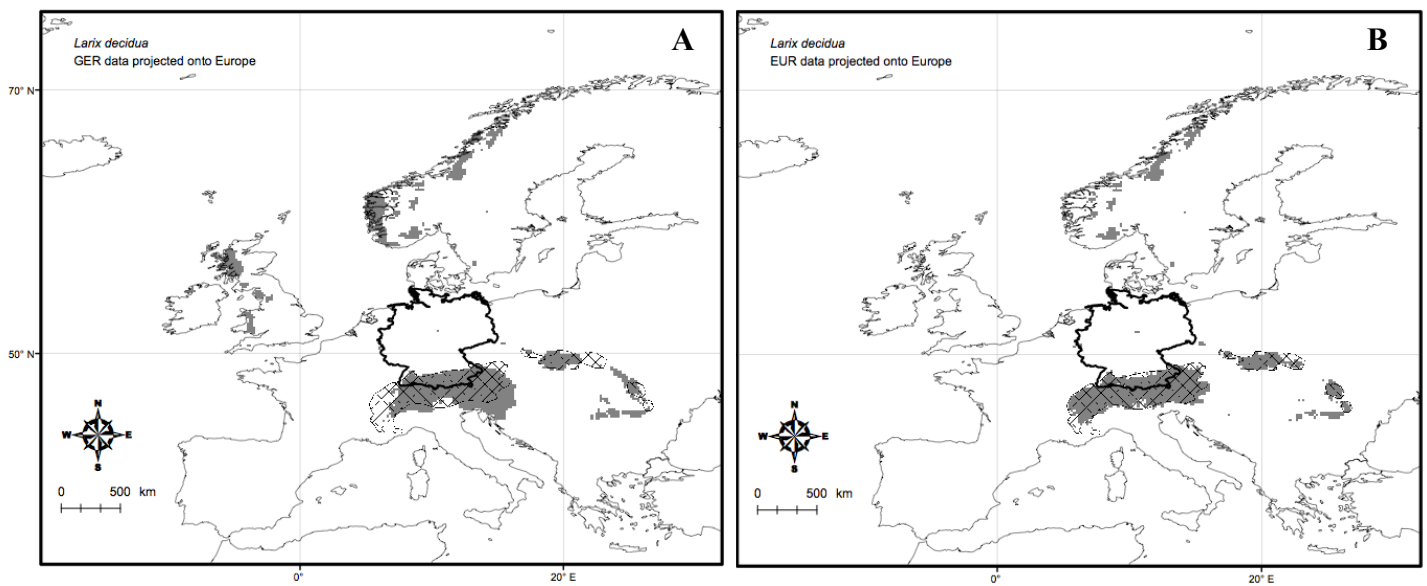


Figure 1 Predicted distributions over Europe (grey) for *Larix decidua* yielded by modelling according to the two distinct training data sets of Germany-only (GER) (A) and full European area (EUR; see Data and Methods) (B). A bold black line highlights the political boundaries of Germany. The probability of occurrence was converted to binary presence/absence by choosing a threshold which optimises the model's AUC statistic. A highlights a small degree of overpredictions in case of truncated training data (GER) being extrapolated on continental scale. B shows the predicted distribution if a more complete spatial data set is used (EUR) for continental scale predictions. The observed current occurrence of *L. decidua* is indicated by the crosshatched area.

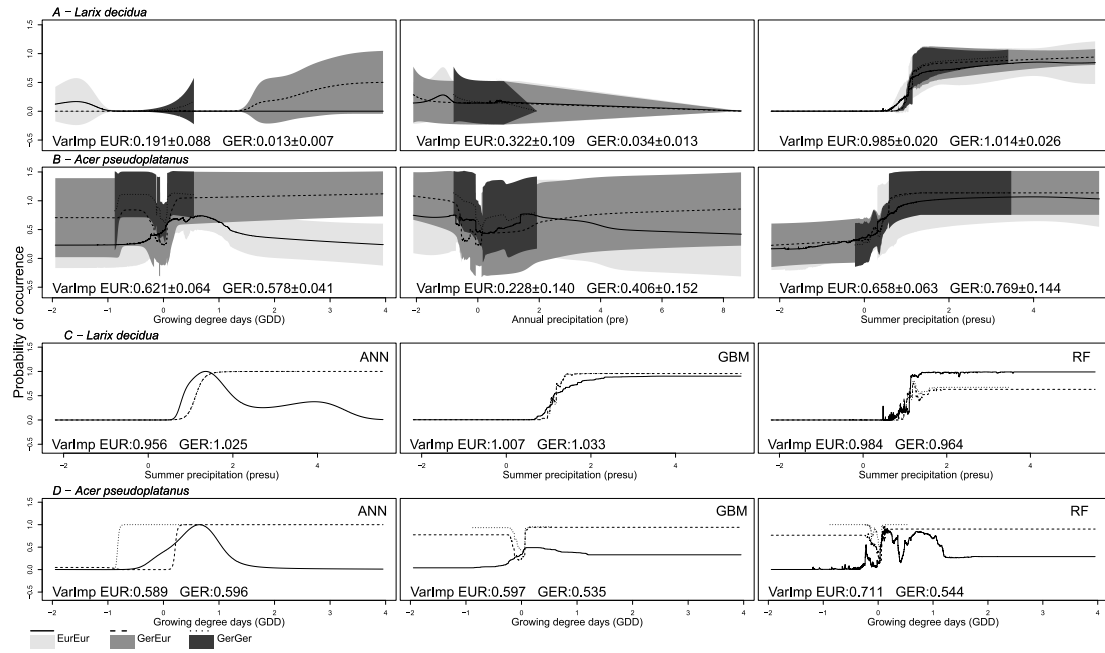


Figure 2 Ensemble and individual modelled response functions for *Larix decidua* and *Acer pseudoplatanus*. GER denotes Germany-only training data whereas EUR indicates training data over the full European area. For all figures the continuous line indicates EUR training data and prediction (EurEur), the dashed line shows GER training data and European prediction (GerEur), and the dotted line indicates GER training data and Germany prediction (GerGer). The variable importance (VarImp) is shown at the bottom of each panel. All variable units have been standardised. **A** shows the ensemble response functions for *L. decidua* for all three variables included in the models, their standard deviation (shaded areas: light grey for EUR predicted on EUR, dark grey for GER predicted on GER, grey for GER predicted on EUR), and the variable importance according to the training data used. **B** shows the same for *A. pseudoplatanus*. Note that in the latter case, large standard deviations and evenly spread values of averaged variable importance lead to over-fitting and largely ecologically unsound model predictions. **C** shows three individual algorithm response functions (artificial neural network - ANN, generalised boosted model - GBM, and random forest - RF) for summer precipitation for *L. decidua*, highlighting the overall

635 consistency of function shapes and high variable weighting leading to consistent
636 model predictions. **D** indicates the equivalent response functions for growing degree
637 days for *A. pseudoplatanus*, showcasing erratic behaviour and irregular variable
638 importance.
639

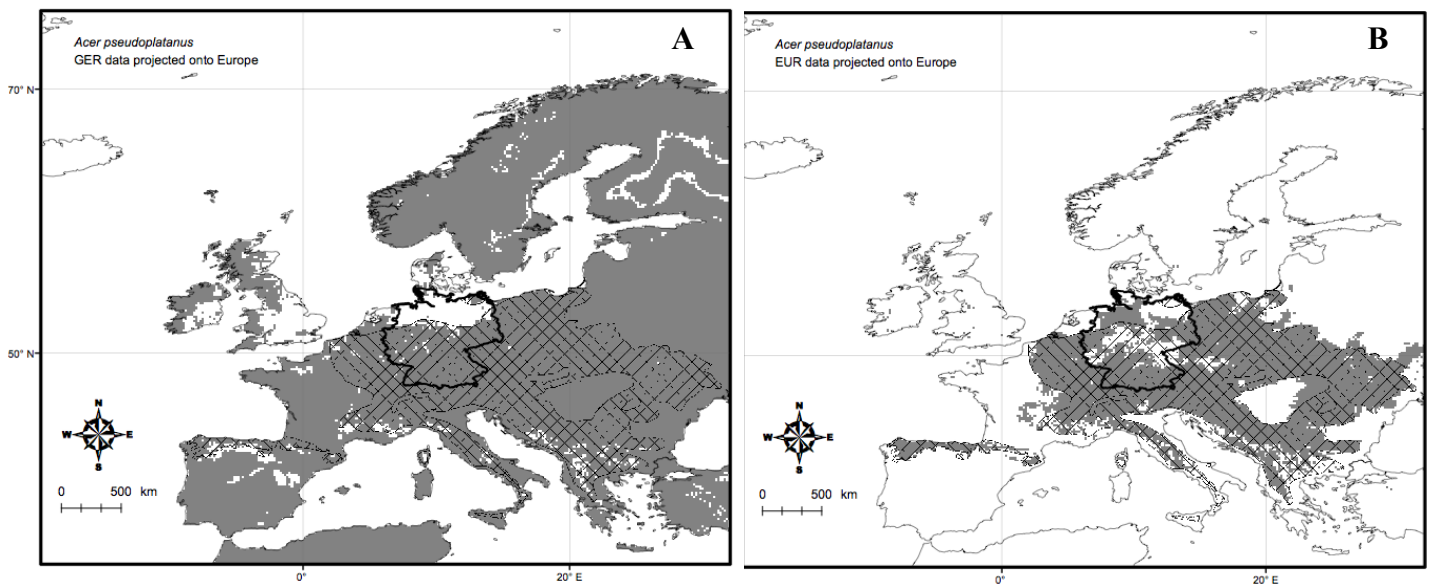


Figure 3 Predicted distributions over Europe (grey) for *Acer pseudoplatanus* yielded by modelling according to the two distinct training data sets Germany-only (GER) (A) and full European area (EUR; see Data and Methods) (B). A bold black line highlights the political boundaries of Germany. The probability of occurrence was converted to binary presence/absence by choosing a threshold which optimises the model's AUC statistic. A highlights the vast overprediction in case of truncated training data (GER) being extrapolated on continental scale. B shows the predicted distribution if a more complete spatial data set is used (EUR) for continental scale predictions. Both maps indicate the observed current occurrence of *A. pseudoplatanus* as crosshatched.