Appendix: ODMAP applications

S1) Bouchet & Meeuwig (2015): Pelagic fishes and sharks

Bouchet & Meeuwig (2015). Drifting baited stereo-videography: A novel sampling tool for surveying pelagic wildlife in offshore marine reserves. Ecosphere, 6: art137.

ODMAP element	**Contents**
OVERVIEW	
Authorship	• **Authors:** Phil J. Bouchet, Jessica J. Meeuwig • **Contact email:** pjbouchet@gmail.com • **Title:** Drifting baited stereo-videography: A novel sampling tool for surveying pelagic wildlife in offshore marine reserves • **DOI:** 10.1890/ES14-00380.1
Model objective	• **Objective:** Mapping/interpolation. • **Target outputs:** Maps of relative probability of presence
Taxon	Pelagic fishes, sharks, and other marine vertebrates
Location	Perth Canyon, Western Australia, Australia.
Scale of analysis	• **Spatial extent (Lon/Lat):** Longitude 115° E – 115.25° E, Latitude 31.90° S – 32.04° S • **Spatial resolution:** 350 m • **Temporal extent/time period**: Autumn (March to May) 2013 • **Type of extent boundary**: Rectangular
Biodiversity data overview	• **Observation type:** Field survey • **Response/Data type:** Presence-only
Type of predictors	Topographic, climatic
Conceptual model / Hypotheses	• **Hypotheses about species-environment relationships:** There is increasing evidence that complex seabed features influence the distribution of marine vertebrates. We used measures of seabed depth, curvature, slope, rugosity, and topographic complexity as predictor variables for fishes and sharks. We also tested sea surface temperature (and its variance).
Assumptions	**We assumed that:** • Relevant ecological drivers (or proxies) of species distributions are included. • Detectability does not change across transects or habitat gradients. • Species are at (pseudo-) equilibrium with their environment. • Species show largely similar responses to the bait and the sampling instrument. Sampling is adequate and representative (and any biases are accounted for/corrected).
SDM algorithms	• **Algorithms**: We fitted MaxEnt and single-visit site-occupancy models to the field data. MaxEnt was chosen due to competitive performance on small

	sample sizes and ease of use, and its outputs were approximated to true probabilities using published equations. Occupancy models were used as a point of comparison, and in an attempt to account for imperfect detectability. (here, detectability allowed to be affected by survey effort, expressed as the log of transect lengths in each grid cell). • **Model complexity**: MaxEnt models were built with linear and quadratic features only. Occupancy models only included linear terms (no interactions). • **Ensembles**: We combined both model types to form ensemble predictions,
Model workflow	Only weakly correlated predictors were retained in the analysis. We performed parameter tuning for the regularisation parameter (b) in MaxEnt models based on AICc, using the ENMTools package. We tested b values from 1 to 10 (in increments of one). Ensemble predictions were derived using weighted means, with weights given by the models' average True Skill Statistic (TSS) evaluated over 50 Bootstrap replicates of the data.
Software	• **Software**: Analyses were conducted in R version 3.1.1 (R Core Team, 2016) and MaxEnt v3.3.3 (http://www.cs.princeton.edu/~schapire/maxent/). • **Data availability**: The raw data are available as a supplementary file accompanying the article.
DATA	
Biodiversity data	• **Taxon names:** All species are listed in the Supplementary Information. • **Ecological level:** Community level. • **Data source**: Survey data collected in the field over a period of 5 days between March and May 2013. • **Sampling design**: Sampling sites were chosen according to a generalized random tessellation stratified (GRTS) spanning a depth range of 1,220 m. Sites were visited once, with six camera units deployed at each location. • **Sample size**: The data contain 55 detections made over 51 camera transects. • **Regional mask:** We clipped all data to the boundary of the study region. • **Scaling:** Sightings were spatially thinned (within 300 m). • **Data cleaning/filtering**: Sightings were lumped across species. • **Background data**: We generated 1,000 random background points within the study area using a bias grid that reflected survey intensity throughout the region (expressed as the log of the cumulative length of transects traversed, in km). • **Errors and biases**: Sighting locations were identified with high precision from time-stamped GPS tracks. Misidentification rates were deemed low, as visibility was generally good and species identifications were checked by three independent image analysts.
Data partitioning	No data partitioning was applied; model performance was assessed using bootstrap resampling (n = 50 replicates).
Predictor variables	• **Predictor variables:** ○ *Topography*: Depth, aspect easting, aspect northing, longitudinal curvature, cross-sectional curvature, slope, slope variance, topographic position index (TPI), fractal dimension, rugosity. ○ *Climate*: Sea surface temperature (SST) and its variance.

	• **Data sources:** Predictors were derived from a national bathymetric grid produced and curated by Geoscience Australia. Remote-sensed sea surface temperature data were obtained from NASA's Multi-scale Ultra-high Resolution (MUR). All data are freely available. URL: http://marine.ga.gov.au/ and https://podaac.jpl.nasa.gov/Multi-scale_Ultra-high_Resolution_MUR-SST • **Spatial resolution and extant of raw data:** The bathymetric grid was available at 350 m resolution. The remote-sensed temperature data were originally available at 1 km resolution. • **Geographic projection:** Asia South Equidistant Conic. • **Temporal resolution and extent of raw data:** SST data were obtained at daily resolution and averaged over the season (March to June). • **Data processing:** SST layers were downscaled to 350 m using bilinear interpolation.
MODEL	
Variable pre-selection	The choice of initial covariates was made as a compromise between their availability and their ecological relevance as indirect proxy of species distributions. Only weakly correlated covariates were included in each model.
Multicollinearity	Multicollinearity between predictors was investigated using variance inflation factors and Spearman rank correlation coefficients. When variables were strongly related (|rs| > 0.5 and/or VIF > 2), we only retained one from each pair to minimize the possibility of over-fitting.
Model settings	• **Model settings:** Default settings for MaxEnt (but tuned regularisation parameter) and single-visit site-occupancy models (linear terms). • **Model settings for extrapolation:** Predictions bounded by clamping in MaxEnt.
Model estimates	Covariate importance was calculated as the sum of model weights for models containing the covariate. Model-averaged parameter estimates from occupancy models are shown in Table 2.
Model averaging / Ensembles	We took an information-theoretic approach to model averaging for each model class, considering all combinations of covariates. Within model classes, candidate models were ranked based on their AICc scores, with models within three units of lowest-AICc model retained and averaged according to their corresponding Akaike weights. Consensus predictions were then obtained across model classes using weighted ensemble means.
Non-independence	Spatial autocorrelation was assessed using spline correlograms in the R package "ncf" (Bjornstad, 2016).
ASSESSMENT	
Performance statistics	• **Performance statistics estimated on training data:** Model performance was assessed based on the average true skill statistic (TSS) from 50 bootstrap runs, calculated based on a TSS-maximisation threshold.
Plausibility checks	• **Response plots:** We used partial dependence plots to check the ecological plausibility of fitted relationships in MaxEnt models.

PREDICTION	
Prediction output	Predictions of relative probability of presence expressed on a continuous scale.
Uncertainty quantification	We compared model outputs to assess algorithmic uncertainty, and applied an ensemble approach averaging over two different model algorithms.

S2) Dormann et al. (2008): SDMs for Great Grey Shrike

Dormann, C. F., Purschke, O., García Marquéz, J. R., Lautenbach, S., & Schröder, B. (2008). Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. Ecology, 89(12), 3371–3386.

ODMAP element	**Contents**
OVERVIEW	
Authorship	• **Authors**: Carsten F. Dormann, Oliver Purschke, Jaime García Márquez, Sven Lautenbach, Boris Schröder • **Contact email**: carsten.dormann@biom.uni-freiburg.de • **Title**: Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike • **DOI**: 10.1890/07-1772.1
Model objective	• **SDM objective**: forecasts/transfer. We quantified uncertainty introduced in various steps of an SDM analysis, from data to climate change scenarios. • **Target output**: probability of occurrence of the target bird species
Taxon	• Great Grey Shrike, *Lanius excubitor*, Laniidae, Passeriformes, Aves
Location	• Saxony, Germany, Europe
Scale of analysis	• **Spatial extent (Lon/Lat)**: 11.87 -15.04E, 50.17 - 51.68, covering 18,416 km$_2$ • **Spatial Resolution:** 5.6 km x 5.6 km • **Temporal extent/time period**: single time slice, 1993-1996 • **Type of extent boundary:** administrative (the German state of Saxony)
Biodiversity data overview	• **Observation type**: standardised monitoring • **Response/Data type**: presence/absence; absences are more likely to be false absences than presences to be false presences
Type of predictors	land cover and climate, soil type, topographic slope and distances to rivers and settlements
Conceptual model / Hypotheses	**Hypotheses about species-environment relationship**: shrike abundance is strongly affected by landscape composition, requiring a mix of hedges for perching and open grassland for hunting; the species is widespread, if not abundant, in Eurasia, without obvious associations to climate
Assumptions	Implicitly, we assume that sampling is homogeneous or related to distance to settlements (which is one of our predictors); for projections, we assume that the current distribution's delimiting factors will also form the niche of the species in the future; that we have included all species-relevant predictors that are likely to change in the future

SDM algorithms	• **Modelling algorithms**: generalised linear model (GLM), generalised additive model (GAM), artificial neural network (ANN) • **Justification of model complexity**: Non-linear and interacting predictors to allow for more realistic cause-effect representation. Model selection to reduce predictive variance; virtual species analysis and randomisation of data analysis to explore the probability of creating spurious results due to the flexible algorithms. • **Model averaging/ensemble modelling used?** No. We explicitly quantified uncertainty introduced in various steps of an SDM analysis.
Model workflow	We tried to always use three variations for every step of the analysis: three data qualities (b, c, d; see Data section), three collinearity methods (none, PCA, sequential regression), three variable selection approaches (stepAIC, best subset AIC, AIC with >20 records per variable), three SDM algorithms (GLM, GAM and ANN), three climate scenarios (A1B, A2 and B2) and three rainfall realisations (wet, normal, dry), as to directly be able to compare their contribution to the overall prediction uncertainty. We combined all levels of the 4 factors of our analysis, yielding 81 models; each model was 5-fold cross-validated (randomly selected); mean CV-AUC is the measure of model quality; each model then predicts to each of the 9 scenario realisations, yielding 729 projections of probability of occurrence for each cell. The workflow is also depicted in Fig. 2 in the paper.
Software	• **Modelling platform**: R (version 2.3.1-2.6.0) with packages MASS, mgcv, nnet, verification (and others); own functions for stepwise selection of ANN; sequential regression; AICcEPV20 • **Code**: code not shared, available on request • **Data**: data not shared, available on request
DATA	
Biodiversity data	• **Taxon names**: single species: *Lanius excubitor* • **Details on taxonomic reference system**: ? (standard bird book, wikipedia) • **Ecological level**: population (depending on personal preferences, this is an arbitrarily defined subpopulation of the species; or it comprises several subpopulations of the Saxonian population) • **Biodiversity data source**: data provided by the Saxonian breeding bird survey (www.umwelt.sachsen.de/de/wu/umwelt/lfug/), orchestrated by the Saxonian Ministry for the Environment; accessed in 2006 • **Sampling design**: specific sampling design and sampling effort unknown (> 400 observers); 550 cells • **Sample size per taxon**: three levels of decreasing confidence of breeding status: prevalence 57/550, 89/550 and 128/550 for d, c and b records, respectively • **Country/region**: Saxony (Germany), 150-1215 m a.s.l. • **Absence data collection**: non-observation of species was treated as absence • **Details on potential errors and biases in data**: breeding records come in three levels of increasing certainty (b to d, e.g. from nest building to fledgling feeding) ; misidentification error low (as visual identification is required for b/c/d classification); bird is conspicuous in behaviour and appearance
Data partitioning	• **Selection of training data:** 5-fold cross-validation (same folds for all approaches) • **Selection of validation data**: (see training data) • **Selection of truly independent test data**: none; we used virtual species analysis and randomisation of data analysis, to explore the probability of creating spurious

	results due to the flexible algorithms, but no external validation
Predictor variables	• **State predictor variables used**: we selected 12 predictors, based on our understanding of the ecology of the bird, but not considering their collinearity (which is part of the model setup): 5 land cover, 1 biotope diversity, slope, distance to rivers, distance to settlements, percentage of sandy soil, mean annual temperature, annual precipitation • **Details on data sources**: web-references to data sources given in the paper footnotes; only land use had to be purchased, all others are freely available; • **Spatial resolution and spatial extent of raw data**: same as biodiversity data • **Geographical projection**: Bessel, Potsdam in Gauss-Krüger (zone 3) coordinates • **Temporal resolution and temporal extent of raw data**: single slice, extent is 4 years (1993-1996) • **Details on data processing and scaling**: climate data were interpolated between climate stations using external drift kriging with elevation as covariate; all variables were transformed aiming at uniformity (typically square root or log-transforms; details see paper) and standardised before further analyses • **Details on measurements errors and bias**: not available are data on prey availability, on sampling effort, on land management, on human disturbance
Transfer data for projection	• **Models and scenarios, and data sources**: ECHAM5 simulations for 2041-2050 with regional downscaling by WETTREG (http://www.cec-potsdam.de/Produkte/Klima/WettReg/wettreg.html), for scenarios A1B, A2 and B2, with three rainfall realisations each (dry, normal, wet) • **Data processing and scaling**: temperature and rainfall scenarios were transformed using the same transformation as for the original data and the original's standard deviation and mean for standardisation
MODEL	
Variable pre-selection	Ecological pre-selection of variables we deemed important for the species, down to 12 predictors
Multicollinearity	Treatment of multicollinearity was tested as part of the study setup: not addressed, PCA and sequential regression.
Model settings	• **Model settings**: GLM (with quadratic terms but no first-order interactions); GAM (with cubic splines); ANN (feed forward, one hidden layer with 7 nodes, decay=0.03). Weights: we considered (and provide data for) but did NOT use weighting cells by their area (as borders had less than 100%)
Model estimates	• **Assessment of model coefficients: --** • **Details on quantification of uncertainty**: (extensive, as this is the aim of the study) • **Assessment of variable importance**: only for GLM
Model selection/Model averaging/Ensembles	• **Model selection strategy**: stepAIC, best subset AIC, best subset with 20 events per variable • no model averaging, no ensemble
Non-	• none

independence correction	
Threshold selection	• Prevalence threshold to assess changes in spatial distributions
ASSESSMENT	
Performance statistics	• **Performance on validation data**: mean AUC on 5-fold cross-validation
Plausibility check	• **Response plots**: partial plot for the four most important predictors in the best model; no CIs
PREDICTION	
Uncertainty quantification	That was the aim of the study; uncertainty in input data, collinearity, model selection, model algorithm, climate scenario and realisation of rainfall; results given as variability of prediction across all 729 combinations. • **Algorithmic uncertainty**: 3 different algorithms (GLM, GAM, ANN) • **Effect of parameter uncertainty, error propagation**: propagation of uncertainty in response, model selection, collinearity, model algorithm, climate change scenario and rainfall scenario onto fit and prediction • **Uncertainty in scenarios**: 3 climate changes scenarios, with three rainfall scenarios each

S3) Fandos and Telleria (2018): SDMs for European migratory birds

Fandos, G., & Tellería, J. L. (2018). Range compression of migratory passerines in wintering grounds of the Western Mediterranean: conservation prospects. Bird Conservation International, 28(3), 462-474.

ODMAP element	**Contents**
OVERVIEW	
Authorship	• **Authors:** Guillermo Fandos and José Luis Tellería • **Contact email:** gfandos@ucm.es • **Title:** Range compression of migratory passerines in wintering grounds of the Western Mediterranean: conservation prospects • **DOI:** 10.1017/S0959270917000120
Model objective	• **SDM objective:** Mapping/interpolation. Calculate the extent of potential breeding and wintering ranges of studied species. • **Target outputs:** continuous occurrence probabilities and binary maps of potential presence for each species
Taxon	European migratory birds, 10 partial migratory passerines (Passeriformes).
Location	**Western Europe**
Scale of analysis	• **Spatial extent (Lon/Lat):** 28° N - 44° N, 10° W - 10° E • **Spatial resolution:** 5x5 km • **Temporal resolution and extent**: 1925-2011. Breeding period (April to August). Wintering period (December-to February)
Biodiversity data overview	• **Observation type**: citizen science, Ring recoveries • **Response/Data type**: presence only
Type of predictors	Climatic, habitat structure
Conceptual model / Hypotheses	Based on previous studies, we selected climate and vegetation structure as important environmental predictor variables related to the large-scale distribution of birds in breeding and wintering areas. Mean temperatures in breeding (April to August) and wintering (December to February) ranges were selected because small birds avoid cold areas to limit excessive thermoregulatory costs (Calder and King 1974). We also included mean precipitation in breeding and wintering grounds. Precipitation improves primary productivity in wintering grounds after the Mediterranean summer (Nahal 1981), thus increasing the carrying capacity for wintering birds (Carrascal and Palomino 2012, Tellería et al. 2014b). However, in temperate areas, higher precipitation has been related to poor breeding success. Thus, it can be suggested that the moistest sectors will be suboptimal areas for breeding (e.g. Öberg et al. 2015). We also included the annual Normalized Difference Vegetation Index (NDVI) as a surrogate of primary productivity and, according to the strong effect of

	vegetation cover on passerine distribution (Wiens 1989), we included woody cover (Di Miceli et al.2011). Finally, we considered the effect of two different variables related to the seasonal pulses of productivity tracked by migratory birds in breeding and wintering grounds (Wisz et al. 2007,Engler et al. 2014). We used precipitation seasonality as a surrogate for the rain-mediated productive pulses in Mediterranean wintering grounds and temperature seasonality as a surrogate of the productive output in central and northern European summer grounds.
Assumptions	• **#1:** Relevant ecological drivers (or proxies) of species distributions are included. • **#2:** Detectability does not change across habitat gradients. • **#3:** Species are at equilibrium with their environment. • **#4:** Sampling is adequate and representative (and any biases are accounted for/corrected).
SDM algorithms	• **Algorithms**: MaxEnt. Chosen due to competitive performance on small sample sizes and presence only data. • **Model complexity**: We allowed 'l' (linear), 'q' (quadratic), 'p' (product), 'and 'h' (hinge) features • **Model averaging**: Model averaging of 10 replicates
Model workflow	Prior to model building, all predictor variables were standardised and used variance inflation (VIF) analysis to avoid highly correlated variables. We only included the five most important and weakly correlated variables per season. Univariate variable importance for each predictor was assessed with jackknife analyses of the regularised gain with training data. Predictive model performance was assessed using a 10-fold cross-validation.
Software	• **Software:** R (no version specified). MaxEnt (v3.3.3k) • **Data availability:** https://euring.org/
DATA	
Biodiversity data	• **Taxon names**: European migratory birds, 10 partial migratory passerines (Passeriformes): Carduelis cannabina; Erithacus rubecula; Fringilla coelebs; Motacilla alba; Phoenicurus ochruros; Phylloscopus collybita; Sylvia atricapilla; Turdus iliacus; Turdus merula; Turdus philomelos • **Taxonomic reference system:** We follow the taxonomy of EURING • **Ecological level**: species • **Data sources:** https://euring.org/ • **Sampling design**: opportunistic; volunteer-based recording schemes • **Sample size:** Carduelis cannabina (62); Erithacus rubecula (170); Fringilla coelebs (55); Motacilla alba (99); Phoenicurus ochruros (68); Phylloscopus collybita (79); Sylvia atricapilla (125); Turdus iliacus (94); Turdus merula (60); Turdus philomelos (301) • **Mask:** we clipped all data to the political boundary of Europe • **Scaling**: Duplicate records in the same localities (e.g. ringing stations) were removed and spatial autocorrelation was minimised by randomly removing occurrences within 5 km of each other. • **Data filtering**: We used individuals with first-encounter records refer to the initial encounter with an individual when it was first ringed in breeding (April to August) or wintering (December to February). Recoveries in wintering (December to

	February) or breeding (April to August), respectively, comprise all dead (including sick and dying) encounters, live recaptures away from the ringing place (nearly all by qualified ringers), and live resightings away from the ringing place (mainly by ringers and other birdwatchers). • **Background data**: random selection of background points in the minimum convex polygon (convex hull) produced by the full set of ringing records. • **Errors and biases**: Error rates deemed low, as volunteers were highly trained and the individual is identified twice (first encounter and recovery) independently. However, the background selection was weighted using a bias grid reflecting human footprint (HFP), an index of population density, land transformation and road density (Sanderson et al. 2002). We used HFP as a bias grid because the mean human footprint per country is positively related to the density of all ringing recoveries reported per country (r = 0.53, P = 0.036, n = 16), a plain index of the effort applied to control ringed birds (Appendix S1).
Data partitioning	We randomly selected 70% of data for model calibration and 30% for validation of the predictions. The models were evaluated using 10-fold cross-validation
Predictor variables	• **Predictor variables:** For breeding and wintering seasonal maximum, minimum and mean temperature, temperature seasonality, seasonal precipitation, precipitation seasonality, sun radiation, NDVI, bare ground, herbaceous cover and tree cover • **Data sources:** Climate: Climate data were recorded from Worldclim 1.4 (Hijmans et al. 2005) and prepared as ASCII raster maps with a resolution of 5x5 km with the raster package in R (Hijmans and Van Etten 2013, R Development Core Team 2015). Vegetation: Normalized Difference Vegetation Index (NDVI) as a surrogate of primary productivity was obtained from MODIS satellite-based sensor (30” resolution, or ~1 km; Carroll et al. 2004, Pettorelli et al. 2011). Woody, Herbaceous and Bare Ground cover from Vegetation Continuous Fields MOD44B (Di Miceli et al.2011) • **Spatial extent:** 180, 180, 60, 90 (xmin, xmax, ymin, ymax) • **Spatial resolution**: The raw resolution of the climate data was 2.5 minutes. The raw NDVI resolution 1 km, and the raw resolution of the vegetation variables was 250m. • **Projection:** WGS 1984, EPSG:4326 • **Temporal extent**: Climate: 1960-1990 • **Data processing**: For NDVI and habitat cover variables we used a spatial aggregation from 1 km to 5 km (final resolution) by the method bilinear (raster R package; Hijmans and Van Etten 2013) • **Dimension reduction**: predictor variables were standardised and used variance inflation (VIF) analysis to avoid highly correlated variables. We only included the five most important and weakly correlated variables per season.
MODEL	
Variable pre-selection	Hypothesised ecological relevance for species distributions.
Multicollinearity	We used variance inflation (VIF) analysis to avoid highly correlated variables (Marquardt 1970). From a set of 11 variables (seasonal maximum, minimum and

	mean temperature, temperature seasonality, seasonal precipitation, precipitation seasonality, sun radiation, NDVI, bare ground, herbaceous cover and tree cover) we selected five environmental predictors per season (VIF scores under 5, range: 1.15–4.56).
Model settings	MaxEnt: log output, featureSet (Autofeatures: Linear, quadratic, product, thresholdand hinge features), regularizationMultiplierSet (b = 1), convergenceThresholdSet (0.00001), samplingBiasRule (human foot print index), samplingBiasNotes (Supplemental Material 1), Replications (10), Random test percentage (30), Random seed (Yes), Replicate Run Type (Crossvalidation)
Model estimates	Covariate importance calculated with jackknife analyses of the regularised gain with training data, which accounts for dependencies between predictor variables by building two sorts of models: one involving a given predictor by itself, and the other involving all features except for the given predictor.
Non-independence analyses	None
Threshold selection	Maxent continuous maps were converted into binary ‘presence–absence’ maps using the threshold selection method based on maximising the sum of sensitivity and specificity (Manel et al. 2001). This threshold is recommended for conservation purposes as omission errors are avoided and sensitivity favoured (Jiménez-Valverde and Lobo 2007, Liu et al. 2013). We considered areas with a habitat suitability above the threshold as ‘presence’ and those below as ‘absence’.
ASSESSMENT	
Performance statistics	The averaged ROC AUC scores were used as model predictive performance on validation data, following a 10-fold cross-validation procedure with ten replicates.
Plausibility check	Response shapes: Partial dependence plots.
PREDICTION	
Prediction output	**Prediction unit:** We used continuous predictions of occurrence probability per species as well as predicted presence per species that were obtained by binarizing the predicted occurrence probabilities using the threshold selection method based on maximising the sum of sensitivity and specificity

S4) Franklin (1998): SDMs for Californian shrubs

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science **9**:733-748.

ODMAP element	Contents
OVERVIEW	
Authorship	• **Authors:** Janet Franklin • **Contact email:** janet.franklin1@gmail.com • **Title:** Predicting the distribution of shrub species in southern California from climate and terrain-derived variables • **DOI:** 10.2307/3237291
Model objective	• **SDM objective**: mapping/interpolation • **Main target output**: continuous habitat suitability index (probability of species presence)
Taxon	20 shrub species that are commonly found in the chaparral and coastal sage scrub communities of southern California
Location	California's southwest ecoregion
Scale of analyses	• **Spatial extent (Lon/Lat):** Not retrievable anymore. • **Spatial resolution:** 30 m • **Temporal extent:** period of climate normals (30 prior years averaged) • **Temporal resolution:** single time period • **Boundary:** natural
Biodiversity data overview	• **Observation type**: field survey • **Response data type**: presence/absence
Type of predictors	climatic, edaphic, topographic
Conceptual Model / Hypotheses	Hypotheses: Climate, soils and topography are correlated with plant distributions
Assumptions	Species–environment equilibrium, Availability of all important predictors
SDM algorithms	• **Model algorithms**: glm, gam, CART • **Model complexity**: we let the data determine model complexity in CT and GAM, and used the response shape of GAMs as indication to define the complexity of GLMs • **Model averaging**: None.

Model workflow	GAMs were first parameterized using all environmental and spatial variables, and smoothing splines, to assess the validity of the linear model for each bivariate relationship, and to explore the shape of the response curve and strength of the relationship between each species and explanatory variable. Then, GLMs were parameterized using all variables and using response functions suggested by the GAM.
Software	• **Software:** S-Plus software version 3.3 for Windows • **Code availability:** not available • **Data availability:** not available
DATA	
Biodiversity data	• **Taxon names:** Adenostoma fasciculatum, Adenostoma sparsifolium, Arctostaphylos glandulosa, Arctostaphylos glauca, Ceanothus crassifolius, Ceanothus greggii-perplexans, Ceanothus leucodermis, Ceanothus tomentosus, Cercocarpus betuloides, Heteromeles arbutifolia, Quercus berberidifolia, Quercus wislizenii, Rhus ovata, Xylococcus bicolor, Yucca whipplei, Artemisia californica, Eriogonum fasciculatum, Malosma laurina, Salvia apiana Salvia mellifera • **Taxonomic reference system**: Jepson Flora of California • **Ecological level**: species • **Data sources**: USDA Forest Service • **Sampling design**: Subjective spatial sampling (representative sampling of plant communities; once-off temporal sampling • **Sample size**: Number of observations 322 to 906 per species, prevalence 25 to 859 presences per species • **Clipping**: National Forests in California Southwest Ecoregion • **Absence data**: plant community plot data yielded presence and absence. Cover data simplified to presence • **Potential errors and biases**: Subjective sampling locations
Data Partitioning	All data were used for model training and model performance was only assessed internally.
Predictor variables	• **Predictor variables**: annual precipitation, mean minimum temperature of the coldest month, mean maxi- mum temperature of the warmest month, Quarterly estimates of Potential solar, radiation, slope, geographical coordinates • **Data source**: J. Michaelsen (unpubl.) provided 1-km gridded estimates of 28 climate variables for the southwest ecoregion • **Data processing**: ○ mean minimum and maximum monthly temperature for each month, and annual and trimesterly mean precipitation were interpolated from climate station data using elevation from a digital elevation model. Universal kriging was used as interpolation method. Interpolated values of the climate variables at the vegetation plot locations are estimated with error (0.5 - 2.0 °C for the temperature variables). ○ Potential solar radiation (insolation) was estimated using the Atmospheric and Topographic Model (ATM) of solar radiation. Topographic variables (elevation, slope and aspect) used in radiation modeling were derived from USGS 7.5’ 30 m × 30 m resolution DEMs. Incoming solar radiation was

	calculated for each grid cell in the DEM, for a single day each month, and then multiplied by the number of days as an estimate of monthly potential radiation. These values were summed to estimate quarterly and annual solar radiation. Errors commonly found in DEMs (related to the method used to derive the elevation grid) are amplified when first- and second-order difference operations are applied to derive slope, aspect. The only additional variable that was tested was slope which was calculated using the IPW software and a second order finite difference method. • **Spatial resolution of raw data**: 1 km, 30 m • **Projection**: UTM
MODEL	
Variable pre-selection	Because there is a body of literature and theory linking bioclimatic gradients to plant distributions, and because the models are intended to be predictive rather than exploratory, explanatory variables were chosen *a priori*.
Multicollinearity	I excluded highly correlated variables based on ecological plausibility. For example, the deviation of the coldest temperature from the annual mean was also examined, but was strongly correlated with jul.maxt. Also, quarterly estimates of solar radiation for the growing season (spring: Mar-May), and driest quarter (summer: Jun-Aug) were highly correlated ($r = 0.88$), and therefore only winter (Dec-Feb) and spring values were used in modelling.
Model setting	I used species presence/absence as the dependent variable, and so the logistic link function and binomial error term were used in the GAMs and GLMs, and classification trees (CT) were developed to predict species presence or absence. GLM response functions were modelled as either linear, quadratic (approximating a unimodal or symmetric Gaussian response curve), or a higher-order polynomial. Classification trees were developed for each species. A cut-off of five observations per terminal node was used, beyond which no splitting of the dataset occurred. Cross-validation was used to determine the optimum tree size yielding the most robust predictions. All trees were pruned to the 22 'best' nodes based on a cost-complexity measure (deleting those 'branches' that reduce deviance the least). Then the tree was pruned inter- actively to the smallest size possible without increasing model error.
Model estimates	• **Uncertainty in model coefficients**: Cross-validation was used to determine the optimum tree size yielding the most robust predictions • **Variable importance**: assessed by quantifying how much deviance each variable explains
Non-independence	• **Method for addressing spatial autocorrelation**: Model residuals were not explicitly tested for spatial dependence. I developed models both with and without geographic coordinates as explanatory variables. This addresses geographic patterning but not spatial dependence or clustering in the dataset. Because of strong clustering of some species among the sample plots, spatial variables accounted for large proportions of the explained deviance in some GAMs and GLMs. However, CT-models that included spatial variables only accounted for a small increase in prediction accuracy over those that did not.
Threshold	I chose an optimum threshold value of p that minimizes omission and commission

selection	errors. This is important because when a species is rare in the sample, a ‘model’ that predicted it to be absent everywhere would have high overall accuracy (but high omission error), and vice versa.
ASSESSMENT	
Performance statistics	• **Performance on training data**: True negative rate, True positive rate, False negative rate, False positive rate. Performance statistics were only estimated on training data. The model fit and significance of the variables were evaluated using the residual deviance. [Austin et al.] suggest deleting observations that are beyond the last positive observation by greater than 1% of the sample, sorted by an important environmental variable. I included the last positive observation +/–10% of the sample, and so the fit of my models may still be overestimated somewhat. The models were also assessed by examining the number of observations (presence/absence) correctly predicted by each model, as well as the proportion of omission (type I) and commission (type II) errors – predicting the species to be absent when it is present, and *vice versa*.
Plausibility check	Response curves were checked for complexity and plausibility. I assessed response shape (linear, unimodal), confidence intervals and optima.
PREDICTION	
Prediction output	• **Prediction unit:** presence and absence • **Post-processing:** I took care to delimit prediction to interpolation within the study area (no prediction or projection to new place or time).

S5) Leitão et al. (2010): SDMs for steppe / farmland birds

Leitão, Moreira & Osborne (2010). Breeding habitat selection by steppe birds in Castro Verde: A remote sensing and advanced statistics approach. Ardeola, 57(E): 93-116.

ODMAP element	Contents
OVERVIEW	
Authorship	• **Authors:** Pedro J. Leitão, Francisco Moreira, Patrick E. Osborne • **Contact email:** p.leitao@geo.hu-berlin.de • **Title:** Breeding habitat selection by steppe birds in Castro Verde: A remote sensing and advanced statistics approach • **DOI:** N/A
Model Objective	• **Objective:** Inference/Explanation**.** • **Target output:** Habitat suitability (relative probability of occurrence), partial dependency plots.
Taxon	Steppe / farmland birds, 15 species
Location	**Baixo Alentejo**, Portugal
Scale of analysis	• **Spatial extent (Lon/Lat):** ○ Baixo Alentejo (coarse scale): Longitude 6.55° W - 8.28° W, Latitude 37.18° N – 38.20° N ○ Special Protection Area for birds of Castro Verde (fine scale): Longitude 7.49° W - 8.14° W, Latitude 37.34° N – 37.53° N • **Spatial resolution:** 1 km grid squares (coarse scale) and 125 m buffer point counts (fine scale) • **Temporal resolution and extent:** Spring (March to May) of 2004 (coarse scale) and 2006 (fine scale) • **Type of extent boundary:** political (boundary of NUTS3 territorial region of Baixo Alentejo for coarse scale; boundary of Special Protection Area for birds of Castro Verde for fine scale)
Biodiversity data overview	• **Observation type:** Visual and auditory detections • **Response/Data type:** Presence/absence data
Type of predictors	• Vegetation, topographic, disturbances
Conceptual model / Hypotheses	We used vegetation / land cover, terrain and landscape disturbances as environmental predictor variables for bird occurrences in our study area.
Assumptions	We assumed that the relevant ecological drivers (or proxies) of species distributions are included, that detectability does not change across transects

	or habitat gradients, that species are at equilibrium with their environment, and that the sampling is adequate and representative (except for one species, which is discussed).
SDM algorithms	• **Algorithms:** SDMs were fitted using Multivariate Adaptive Regression Splines (MARS) • **Model complexity:** Model settings, relating to the use of interactions between variables and model backfitting penalisation, were chosen by grid search optimising for model performance. • **Model averaging:** We fitted models with different subsets of predictor variables, based on different correlation thresholds. For inference of the species habitat preferences (at each scale of analysis), we considered only the responses which were consistently fitted across correlation thresholds (i.e. those that entered all models), and calculated their averaged model drop contributions (explained deviance loss when variable dropped).
Model workflow	We used a cross-correlation approach, by running three models per species (at each scale), with different subsets of predictor variables, based on different data dependency levels. We hence defined three thresholds of (rank) correlation between the variables – at Spearman rho values of 0.7, 0.6 and 0.5 – for all species, on both 'fine' and 'coarse' models. Using these defined thresholds, we generated three predictor subsets, from the original sets. In each case, the best fitting set of uncorrelated variables (below the respective rho value) was kept, the remaining variables being discarded. The averaged ROC AUC scores were used as model predictive performance measures, following a 10-fold cross-validation procedure with five replicates.
Software	• **Software:** All analyses were conducted using R (no version specified; R Core Team, 2010) with a modified version of code from the 'mda' package, to allow for binary data and n-fold model cross-validation (Hastie and Tibshirani, 1996; Elith and Leathwick, 2007) • **Code availability:** the code used was kindly provided by Jane Elith (pers.comm.), and were later published in package 'earth'. • **Data availability:** no data were made available.
DATA	
Biodiversity data	• **Taxon names:** All species are listed in Table 1 of the original publication. • **Taxonomic reference system:** not specified. • **Ecological level:** population level. • **Data source:** Bird presence-absence data were collected in the field by trained surveyors. • **Sampling design:** Bird atlas data (coarse scale) were recorded during Spring 2004, according to a geographically stratified random scheme; bird point counts (fine scale) were collected during Spring 2006, following a combined systematic and random scheme. In both cases, species were identified by visual or auditory cues, down to the species level (except for two species which were considered together due to the non-reliable field discrimination). • **Sample size:** 557 grid squares (coarse scale) and 1076 point counts (fine scale). • **Absence data:** Count data converted to presence/absence. Zero count was interpreted as absence.

Data partitioning	10-fold cross-validation
Predictor variables	• **Predictor variables:** o *Vegetation:* seven variables which describe the yearly vegetation phenological changes (as measured by NDVI): three relating to vegetation vigour changes, such as senescence and growth in different periods (Summer, Winter and Spring); two relating to average vegetation vigour or biomass in peak periods (Low and High); and two single month NDVI averages, which were uncorrelated with all previous measures (Dec and May) for the coarse scale study; and four land cover classes - Fallow, Cereal, Bare soil and Woodlands & shrubs - as well as one class related with the phenological gradient of herbaceous vegetation (Phenology), ranging from green to dry (an indicator of soil water availability) for the fine scale study. o *Terrain:* Altitude and topographic variability (at a 10m vertical resolution) for the coarse scale study; and Slope and 'Terrain variability (standard deviation of slope) for the fine scale study. o *Disturbance:* Distance to the nearest urban or constructed structures, to the nearest paved road and to the nearest river, for the coarse scale study; and distance to the nearest water body, to the nearest paved road, to the nearest built-up structure, to the nearest tree, and tree density, for the fine scale study. • **Data source:** The coarse scale vegetation predictors were derived from a time series of Normalized Difference Vegetation Index (NDVI) imagery from the Spot Vegetation sensor. The fine scale vegetation predictors, as well as the distance to the nearest water body, were derived from a two-image set of Landsat TM data. The coarse scale terrain predictors were derived from a DTM acquired from the Instituto Geográfico Português (IGP). The fine scale terrain predictors were derived from airborne laser scanning (LiDAR) data. The coarse scale disturbance predictors were derived from Corine Land Cover 2000 raster data (European Environment Agency), a vector-based road map provided by the Instituto de Estradas de Portugal (IEP) and a hydrographical map from the Agência Portuguesa do Ambiente – Atlas do Ambiente Digital (http://www2.apambiente.pt/atlas/est/index.jsp). At the fine scale, the disturbance predictors were derived from airborne laser scanning (LiDAR) data and a road map compiled using a Global Positioning System (GPS). • **Data processing:** Both direct (raw) and indirect (processed) remote sensing data products were used to describe environmental and landscape features related to vegetation, terrain and disturbance. o *Landsat:* At the fine scale, the main land cover classes in the Castro Verde area were characterised by classifying a two-image set of Landsat TM data ("Fallow", "Cereal", "Bare soil" and "Woodlands & shrubs"), as well as one class related with the phenological gradient of herbaceous vegetation ("Phenology"), ranging from green to dry (senescent) – an indicator of soil water availability. From this classification we extracted a map of all water bodies in the region, from which we derived the variable "Waterdist", distance to the nearest water body. This variable constitutes

	a fragmenting element in the pseudo-steppe landscape, and was thus grouped together with the disturbance variables. ○ *LiDAR:* Airborne laser scanning (LiDAR) data were collected over the Castro Verde study area. Variables derived from these data were terrain slope (“Slope”), terrain variability (“Terrainvar”), distance to the nearest built-up structure (“Builtdist”) and to the nearest tree (“Treedist”) and tree density (“Treedens”). All roads and tracks in the area were mapped using a Global Positioning System (GPS), and the distance to these calculated in the variable “Roaddist”. • *NDVI:* At the coarse scale, the temporal vegetation patterns in the Baixo Alentejo region were described by a 12 month series of Normalized Difference Vegetation Index (NDVI) imagery from the Spot Vegetation sensor. This temporal series refers to the 12 month period preceding the end of the respective field season (June 2003-May 2004). The series was further reduced into seven uncorrelated variables which describe the yearly vegetation phenological changes: three relating to vegetation vigour changes, such as senescence and growth in different periods (“Summer”, “Winter” and “Spring”); two relating to average vegetation vigour or biomass in peak periods (“Low” and “High”); and two single month NDVI averages, which were uncorrelated with all previous measures (“Dec” and “May”). The region’s altitude (“Alt”) and topographic variability at 10 m vertical resolution (“Topov10”) were extracted from a DTM. Corine Land Cover 2000 raster data were used to calculate a distance map to the nearest towns, urban settlements or other constructed structures (“Urbandist”). Distance to the nearest road (“Roaddist”) and nearest river or water body (“Waterdist”) were calculated from vector-based maps.
MODEL	
Variable pre-selection	Compromise between covariate availability and ecological relevance as indirect proxy of species distributions.
Multicollinearity	Collinearity assessed using spearman rank. Cross-correlation approach, by running three models per species (at each scale), with different subsets of predictor variables, based on different data dependency levels. We hence defined three thresholds of (rank) correlation between the variables – at Spearman rho values of 0.7, 0.6 and 0.5 – for all species, on both ‘fine’ and ‘coarse’ models. Using these defined thresholds, we generated three predictor subsets, from the original sets. In each case, the best fitting set of uncorrelated variables (below the respective rho value) was kept, the remaining variables being discarded.
Model settings	Model fine-tuning, relating to the use of interactions between variables and model backfitting penalisation, was done by grid search optimising for model performance.
Model estimates	Covariate importance calculated as averaged model drop contributions (explained deviance loss when variable dropped)

Model averaging / ensembles	For inference of the species habitat preferences (at each scale of analysis), we considered only the responses which were consistently fitted across correlation thresholds (i.e. those that entered all models), and calculated their averaged model drop contributions (explained deviance loss when variable dropped).
Non-independence analyses	N/A
ASSESSMENT	
Performance statistics	• **Performance statistics estimated on validation data (from data partitioning):** The averaged ROC AUC scores were used as model predictive performance measures, following a 10-fold cross-validation procedure with five replicates.
Plausibility check	The fitted species responses were checked for ecological plausibility through the inspection of partial dependency plots.

S6) Peterson and Samy (2016): Filovirus Distributions

Peterson AT & Samy AM (2016) Geographic potential of disease caused by Ebola and Marburg viruses in Africa. Acta Tropica 162: 114-124.

ODMAP element	**Contents**
OVERVIEW	
Authorship	• **Authors:** A. Townsend Peterson, Abdallah M. Samy • **Contact email:** town@ku.edu • **Title:** Geographic potential of disease caused by Ebola and Marburg viruses in Africa • **DOI:** 10.1016/j.actatropica.2016.06.012
Model objective	• **SDM objective**: Mapping/interpolation. **Main target output**: Potential species presence
Taxon	Ebola and Marburg viruses
Location	Africa
Scale of analysis	• **Spatial extent (lon/lat):** 19° W – 53° E, 36° S – 38° N • **Spatial resolution:** 1 km • **Temporal extent:** Occurrence data extend 1960s to present; environmental data for just one year (1995-1996), approximating the median date of the occurrence data. • **Type of extent boundary:** rectangular
Biodiversity data overview	• **Observation type:** Disease case records • **Response data type:** Presence-only
Type of predictors	Vegetation index
Conceptual model / Hypotheses	• **Hypotheses about species-environment relationships**: Focus was on the ecological niche of each species, but analysis was limited to areas posited as accessible to each species, so the output would be a distribution model, more or less.
Assumptions	All of the assumptions in Table 2 apply, in some sense. Perhaps the most relevant, however, is that the very limited sampling available to this study represents enough of the true niche signal that the estimated niche is informative about the distributional potential of the species.
SDM algorithms	• **SDM algorithm:** Maxent; other algorithms tested initially via the openModeller platform. Choice of Maxent was based on bootstrapping procedures it includes, which facilitate calculation of uncertainty. • **Model complexity:** Not controlled quantitatively. However, principal components analysis was used to control the complexity of the environmental

	space and limit the number of dimensions involved. - **Model averaging/ensembles:** N/A
Model workflow	Model calibration done with careful consideration of data limitations, and effects of various assumptions available were assessed via experimentation. No statistical evaluation was done owing to small sample sizes, but uncertainty was considered carefully.
Software	- **Software:** Maxent 3.3.3k - **Code availability:** N/A - **Data availability:** Data are available in an open, online, digital repository.
DATA	
Biodiversity data	- **Taxon names:** Zaire ebolavirus, Sudan ebolavirus, Marburg - **Taxonomic reference system:** N/A - **Ecological level:** species - **Biodiversity data source:** Data were derived from literature sources, but all data are openly available, and the version is clear because the data matrix is deposited as it was at the time of analysis. Including our earlier compilation (Peterson et al., 2004; Peterson et al., 2006) and numerous subsequent compilations (Bausch and Schwarz, 2014; Changula et al., 2014; Chippaux, 2014; Leroy et al., 2009; Mylne et al., 2014; Pourrut et al., 2005; Roddy, 2014). As an independent source, however, we reviewed all posts in the ProMed archives (http://www.promedmail.org/, queries executed 1 November 2014) that included reference to “Ebola,” “Marburg,” or “filovirus” - **Sampling design:** disease case records - **Sample size per taxon:** 1-34 - **Background data:** Random, as no information on the sampling process was available. - **Details on data cleaning:** Only occurrences cited as “confirmed” were used, and records were always verified against independent information sources. We omitted occurrences detected serologically in bats, in view of the rather odd patterns that such detections have shown. For each occurrence in the final list, we used Internet-based electronic gazetteers to add geographic coordinates to the data record. Occurrence data were random points within the radius of uncertainty around each point. - **Details on potential errors and biases:** We assigned a rough estimate of the uncertainty associated with each data record (entering a cave was assigned 200 m uncertainty; a person in a village was assigned 5–10 km; a general description of a region over which a human was infected was assigned 50–170 km, depending on the density of other major landmarks or political regions).
Data partitioning	We conducted 10 replicate analyses for each species based on a 50% bootstrap of available occurrence data.
Predictor variables	- **Predictor variables**: Normalized Difference Vegetation Index (NDVI) - **Data source:** we used the old imagery for NDVI because known filovirus outbreaks date back as far as 1976. Specifically, data stem from the AVHRR satellite (James and Kalluri, 1994). - **Spatial resolution:** 1 km

	• **Map projection:** WGS 1984 • **Temporal resolution and extent:** We used 12 monthly composite NDVI data layers (downloaded, with atmospheric corrections already completed, from UMD, 2001) corresponding to February 1995–January 1996 to capture aspects of land cover and seasonality. These data layers correspond approximately to the midpoint of the time span of filovirus occurrence data used on model development. • **Dimension reduction:** Details provided - input NDVI layers were subjected to a principal components analysis, to reduce dimensionality and create orthogonal input layers.
MODEL	
Multicollinearity	Principal components analysis was used for dimension reduction prior to modelling. This reduces collinearity greatly or eliminates it entirely.
Model settings	• **Model settings:** Maxent, without model selection approaches. Data were weighted in effect, based on the uncertainty associated with georeferencing, as the random points for highly uncertain points were quite variable.
Model estimates	• **Assessment of model coefficients**: We used the jackknife option to identify variables not contributing importantly to model robustness
Non-independence analyses	None.
Threshold selection	Not used; in light of the small sample sizes available, we generally retained continuous model outputs for interpretation and visualization of suitability patterns.
ASSESSMENT	
Performance statistics	None used in light of very small sample sizes.
Plausibility check	No plausibility checks conducted.
PREDICTION	
Prediction output	• **Prediction unit:** median logistic output of replicate analyses
Uncertainty quantification	• **Uncertainty in input data:** Evaluated carefully in terms of spatial uncertainty in geographic coordinates, as a function of information available about the origin of each filovirus outbreak. Variation among model replicates was used as an index to stability or instability of estimates resulting from modelling process.

S7) Rapacciuolo et al. (2012): SDMs for British birds, butterflies and plants

Rapacciuolo G, Roy DB, Gillings S, Fox R, Walker K, et al. (2012) Climatic Associations of British Species Distributions Show Good Transferability in Time but Low Predictive Accuracy for Range Change. PLOS ONE 7(7): e40212. https://doi.org/10.1371/journal.pone.0040212

ODMAP element	Contents
OVERVIEW	
Authorship	• **Authors:** Giovanni Rapacciuolo, David B. Roy, Simon Gillings, Richard Fox, Kevin Walker, Andy Purvis • **Contact email:** grapacciuolo@calacademy.org • **Title:** Climatic Associations of British Species Distributions Show Good Transferability in Time but Low Predictive Accuracy for Range Change • **DOI:** 10.1371/journal.pone.0040212
Model objective	• **Objective**: Mapping/interpolation and forecast/transfer • **Target outputs**: continuous occurrence probabilities and binary maps of potential presence
Taxon	(i) birds, (ii) butterflies, and (iii) vascular plants
Location	Great Britain
Scale of analysis	• **Spatial extent (Lon/Lat)**: Longitude -7.57°–8.12°, Latitude 49.96°– 58.64° • **Spatial resolution**: British Ordnance Survey 10km grid squares • **Temporal resolution and extent**: Two time periods for birds (t1: 1968–1972; t2: 1995–1999), butterflies (t1: 1970–1982; t2: 1988–1991), and vascular plants (t1: 1930-1969; t2: 1987-1999) • **Type of extent boundary**: political/natural (coastlines Great Britain)
Biodiversity data overview	• **Observation type**: volunteer-based surveys • **Response/Data type**: presence/pseudo-absence data
Type of predictors	Climatic, (topography and geology not included in final models)
Conceptual model / Hypotheses	We tested whether the most commonly-used SDM algorithms were transferable over time by calibrating models using survey data in one time period and validating them against survey data in a second time period.
Assumptions	We assumed that species are at pseudo-equilibrium with the environment. We assumed that each surveyed grid square in which a species was not recorded (i.e., non-detection) represented an absence.
SDM algorithms	• **Model algorithms**: We built SDMs using 7 presence-absence modelling algorithms (classification tree analysis, CTA; generalised linear models, GLMs; generalised additive models, GAMs; multi-variate adaptive regression splines, MARS; artificial neural networks, ANNs; generalised boosted models, GBMs; and

	random forests, RFs), two presence-only modelling algorithms (maximum entropy, MaxEnt and surface range envelope, SRE) and one ensemble method (the mean probability of occurrence from all seven presence-absence modelling algorithms). • **Model complexity**: We chose different modelling parameters to optimise each statistical technique (see Supporting Information, Appendix S1). Model settings were chosen to yield intermediately complex response surfaces but prevent excessive overfitting. • **Model averaging/ensemble**: the seven presence-absence modelling algorithms were combined in an ensemble
Model workflow	• We used the species-climate associations identified in period t1 to generate predictions of each species' geographic distribution in (a) time period t1 (interpolation to the same climate used to build the models) and (b) period t2 (extrapolation to the climate experienced in the more recent period), based on observed climate for the corresponding periods. • We measured the agreement between predictions in period t2– as generated by each of the nine single-models built in period t1 plus the consensus method – and observed presence-absence for the corresponding period using three alternative measures of prediction accuracy: (i) area under curve (AUC) of the receiver operating characteristic (ROC) function, (ii) sensitivity (i.e., proportion of correctly-predicted presences), and (iii) specificity (i.e., proportion of correctly-predicted absences). • We measured the agreement between observations and model predictions of each species' (a) geographic range size in period t2, (b) overall change in geographic range size between time periods, and (c) grid square-level changes in occupancy status between time periods.
Software	• **Software:** All analyses were run in the BIOMOD R package version 1.0 (Thuiller et al. 2009) using R version 2.13.0. • **Data availability:** N/A • **Code availability:** N/A
DATA	

Biodiversity data	• **Taxon names**: The full lists of species names surveyed over all time periods are available at nbnatlas.org
	• **Taxonomic reference system**: We follow the taxonomy of the organizations responsible for each survey; British Ornithology Trust for birds, Butterfly Conservation for butterflies, and the Botanical Society of the British Isles for vascular plants
	• **Ecological level**: population level
	• **Biodiversity data source:**
	  ○ Birds: We used records for 183 bird species in 2808 British 10km grid squares at two time periods (t1: 1968–1972; t2: 1995–1999), corresponding to the periods of intensive recording effort leading to the publication of two national atlases of butterflies (Sharrock 1976; Gibbons, Reid & Chapman 1993). We excluded all species occurring in fewer than 20 10km grid squares across Great Britain.
	  ○ Butterflies: We used records for 53 butterfly species in 2808 British 10km grid squares at two time periods (t1: 1970–1982; t2: 1988–1991), corresponding

	to the periods of intensive recording effort leading to the publication of two national atlases of butterflies (Heath et al. 1984; Asher et al. 2001). We excluded all species occurring in fewer than 20 10km grid squares across Great Britain. - Plants: We used records for 1587 vascular plant species in 2808 British 10 km grid squares at two time periods (t1: 1930-1969; t2: 1987-1999), corresponding to the periods of intensive recording effort leading to the publication of two national atlases of the British flora (Perring & Walters 1962; Preston et al. 2002). We excluded all species occurring in fewer than 20 10km grid squares across Great Britain. - **Sampling design:** Data were collected during volunteer-based surveys aimed at visiting each 10km grid square at least once in each time period. Many grid squares were visited repeatedly during each time period and extra visits were organized to survey under-sampled grid squares. - **Sample size:** The bird dataset contained 183 bird species in 2808 10km grid squares at two time periods (t1: 1968–1972; t2: 1995–1999). The butterfly dataset contained 53 butterfly species in 2808 10km grid squares at two time periods (t1: 1970–1982; t2: 1988–1991). The plant dataset contained 1587 vascular plant species in 2808 British 10 km grid squares at two time periods (t1: 1930-1969; t2: 1987-1999). - **Country mask:** All data were included within the British Ordnance Survey grid across the British Isles. - **Absence data:** Although the absence of species from each 10 km grid square could not be definitively recorded during sampling, most grid squares surveyed in each period (i.e., 92–100% of Great Britain's 10 km grid squares) were meticulously sampled, with high levels of duplicate recording and under-recorded areas being targeted by extra recording schemes. Thus, we assumed that each surveyed grid square in which a species was not recorded (i.e., non-detection) represented an absence. We acknowledge that sampling extent and intensity did vary among surveys and taxonomic groups; we discuss the potential implications of this heterogeneity on results.
Data partitioning	No partitioning of calibration data (i.e. t1) was used, since models were validated using temporally independent data (i.e. t2).

Predictor variables	- **Predictor variables:**
	  - Climate: mean temperature of the coldest month (MTCO, °C), mean temperature of the warmest month (MTWA, °C), ratio of actual to potential evapotranspiration (APET, standard moisture index), potential sunshine (PSUN, minutes), total annual precipitation (TPRE, mm), and the difference between total winter precipitation and total summer precipitation (PREvar, mm)
	  - Topography (tested, but not included in final models): median and standard deviation of elevation, m
	  - Geology (tested, but not included in final models): percentage cover of five substrate classes in each 10 km grid square: igneous and metamorphic; peat; sedimentary acid; sedimentary basic; and superficial
	- **Data sources:** Monthly values of temperature, precipitation and cloud cover for each year between 1930 and 1999 were obtained from the CRU ts2.1 and the CRU 61–90.
	- **Spatial resolution and extent of raw data:** Climate data were interpolated

	from weather stations at the 10km grid of the species data. • **Temporal resolution and extent of raw data:** Climate data were averaged across the time periods corresponding to each species dataset. • **Data processing:** o Monthly climate data were used to calculate mean values for nine climate variables – separately for each t1 and t2 period – that reflect hypothesised physiological constraints on species survival and growth. We also considered including additional environmental predictors of ecological relevance to our models. First, although changes in land use have been identified as fundamental drivers of change for many British species [48–52], we were unable to account for them in our models – like most other published accounts of temporal transferability of SDMs [20,21,24,25] – due to the lack of data documenting habitat use in the earlier t1 period; detailed digitised maps of land use for the whole of Britain are not available until the UK Land Cover Map in 1990 [53]. o Topography: we derived two topographic variables (median and standard deviation of elevation, m) and five geological variables (percentage cover of five substrate classes in each 10 km grid square: igneous and metamorphic; peat; sedimentary acid; sedimentary basic; and superficial). • **Dimension reduction:** We compared the performance between models built using climate predictors vs. Climate and topography predictors. Models including geology and topography predictors as well as climate had a higher accuracy than models with climate only according to most performance measures calculated, including both calibration and validation AUC; however, they had a lower mean correct classification rate for squares having changed occupancy status between time periods. For this reason, we decided to leave both topography and geology variables out from our final models.
Transfer data	To assess temporal transferability in t2, we used records for 183 bird species between 1988-1991 (Gibbons, Reid & Chapman 1993), records for 53 butterfly species between 1995–1999 (Asher et al. 2001) and records for 1587 vascular plant species between 1987-1999 (Preston et al. 2002) across the same 2808 British 10km grid squares used for calibration.
MODEL	
Variable pre-selection	The six climate variables included in all models were mean temperature of the coldest month (MTCO, °C), mean temperature of the warmest month (MTWA, °C), ratio of actual to potential evapotranspiration (APET, standard moisture index), potential sunshine (PSUN, minutes), total annual precipitation (TPRE, mm), and the difference between total winter precipitation and total summer precipitation (PREvar, mm).
Multicollinearity	We conducted Spearman's rank correlations between all pairs of climate variables and dropped three variables that were highly correlated with others (Spearman's $|\rho| < 0.85$) to reduce the risk of overfitting during model calibration. The final six climate variables included in the models were mean temperature of the coldest month (MTCO, °C), mean temperature of the warmest month (MTWA, °C), ratio of actual to potential evapotranspiration (APET, standard moisture index), potential sunshine (PSUN, minutes), total annual precipitation (TPRE, mm), and the difference between total winter precipitation and total summer

	precipitation (PREvar, mm).
Models estimates	We did not analyse model coefficients in depth.
Model settings	• Classification tree analysis (CTA): CTA was carried out using a 10-fold cross-validation to select the best trade-off between the number of leaves on the tree and the explained deviance. • Generalised linear models (GLMs): GLMs were generated assuming a logistic link function and a binomial error distribution of the response. Linear, quadratic and polynomial terms (second and third order) of each climatic predictor were included in the initial models, and a stepwise procedure using the AIC criterion was used to select the most significant terms. • Generalised additive models (GAMs): GAMs were generated with cubic-smooth splines bounded by a degree of smoothness of four for each climatic predictor. A stepwise procedure similar to GLMs was used to select the most parsimonious models. • Multivariate adaptive regression splines (MARS): MARS were fitted with two-level interactions between predictors. • Feed-forward artificial neural networks (ANNs): The accuracy of ANNs is mainly controlled by two parameters: the amount of weight decay and the number of hidden units. In this analysis, these two parameters were set to 0.03 and 7, respectively. Due to their heuristic nature, ANNs were run 10 times for each species and the average prediction used. • Generalised Boosted Models (GBMs): GBMs were fitted with an interaction depth of 4, a learning rate of 0.001, and a maximum of 5000 trees fitted to the data. • Random Forests (RFs): the number of trees grown were set to 500 and the number of predictors to be chosen randomly at each node were set to (total number of predictors – 1) • Surface range envelopes (SREs): absence locations were only included to the species' potential range if falling within the 2.5- and 97.5-percentiles of the species' environmental envelope. • Maximum entropy (MaxEnt): We enabled the use of all six feature classes (linear, product, quadratic, hinge, threshold and categorical) for modelling species responses to environmental variables. The default value of 1.0 was used as the regularization parameter, which affects how closely the training data is fitted.
Model averaging / Ensembles	We calculated the mean probability of occurrence from all seven presence-absence modelling techniques (abbreviated Mn(PA)) as a simple but efficient consensus method for combining the output of different single-models.
Non-independence	Potential non-independence in the data was not accounted for in the models
Threshold selection	We selected an appropriate probability threshold to turn continuous probabilities of occurrence into binary presence-absence predictions. For each model, we calculated the sum of sensitivity and specificity on calibration data (i.e. t1) for 100 threshold values (in 0.01 increments), and selected the threshold that maximized this sum; this threshold has previously been found to perform well in comparisons with others (Jiménez-Valverde & Lobo 2007) .. Predicted

	probabilities of occurrence at time t2 above the selected threshold were converted to presences and those below to absences.
ASSESSMENT	
Performance statistics	• **Performance statistics on test (truly independent) data:** To quantify the transferability of SDMs in time, we measured the agreement between forecasts in period t2– as generated by each of the nine single-models built in period t1 plus the consensus method – and observed presence-absence for the corresponding period using three alternative measures of prediction accuracy (Fielding & Bell 1997): (i) area under curve (AUC) of the receiver operating characteristic (ROC) function, (ii) sensitivity (i.e., proportion of correctly-predicted presences), and (iii) specificity (i.e., proportion of correctly-predicted absences).
Plausibility checks	Maps of modelled predictions were checked by experts for an ad-hoc subset of species.
PREDICTION	
Prediction output	• **Prediction unit:** Quantifying the temporal transferability of SDMs by comparing the agreement between model predictions and observations for the predicted period using common metrics is not a sufficient test of whether models have actually captured relevant predictors of change. A single range-wide measure of prediction accuracy conflates accurately predicting species expansions and contractions to new areas with accurately predicting large parts of the distribution that have remained unchanged in time. Thus, to assess how well SDMs capture drivers of change in species distributions, we measured the agreement between observations and model predictions of each species' (a) geographic range size in period t2, (b) overall change in geographic range size between time periods, and (c) grid square-level changes in occupancy status between time periods. We measured the agreement between observed and predicted range size in t2 and between observed and predicted overall change in range size across time periods using Spearman's r statistic. To calculate the agreement between observed and predicted grid square-level changes in occupancy status, we divided binary forecasts into (a) grid squares that had either remained occupied or remained unoccupied between time periods and (b) grid squares that had changed occupancy status (from occupied to unoccupied or vice versa) between time periods. We then measured the correct classification rate (CCR; i.e., the sum of true positives and true negatives divided by the total number of locations) of grid squares in each of these two subsets for each modelling technique, to capture how well our models predict stable versus dynamic portions of each species' distribution.
Uncertainty quantification	• **Algorithmic uncertainty:** In addition to predictions from these nine single models, we calculated the mean probability of occurrence from all seven presence-absence modelling techniques (abbreviated Mn(PA)) as a simple but efficient consensus method for combining the output of different single-models (Marmion et al. 2009); this approach can reduce model-based uncertainty in predictions from SDMs (Araújo & New 2007).

S8) Schröder et al. (2009): SDMs for two butterfly species

Schröder B, Strauss B, Biedermann R, Binzenhöfer B, Settele J. 2009. in Ecology of butterflies in Europe (Eds. Settele J, Shreeve TG, Konvicka M, Van Dyck H) 62-78, Cambridge University Press, Cambridge.

(re-analysis of Binzenhöfer B, Schröder B, Strauss B, Biedermann R, Settele J. 2005. Habitat models and habitat connectivity analysis for butterflies and burnet moths – The example of *Zygaena carniolica* and *Coenonympha arcania*. Biological Conservation 126, 247-259.).

ODMAP element	Contents
OVERVIEW	
Authorship	• **Authors:** Boris Schröder, Barbara Strauss, Robert Biedermann, Birgit Binzenhöfer, Josef Settele • **Contact email:** boris.schroeder@tu-bs.de • **Title:** Predictive species distribution modelling in butterflies.
Model objective	• **Objective:** Mapping/interpolation**.** The models were used as a case study for illustrating recommendations for species distribution modelling. Originally (Binzenhöfer et al. 2005), they were developed to provide insights into (i) the factors controlling the spatial distribution of the studied species within the study area and (ii) to generate habitat suitability maps as a basis for habitat connectivity analyses. • **Target output:** continuous occurrence probabilities), habitat suitability maps, partial dependence plots.
Taxon	Butterflies, two species.
Location	Nature reserve "Hohe Wann" in Northern Bavaria, Germany
Scale of analysis	• **Spatial extent (Lon/Lat):** Longitude 10°35'E, Latitude 50°03'N • **Spatial resolution:** 118 sample plots in the area, 30 × 30 m each • **Temporal resolution and extent:** data collection in 2001 and 2002
Biodiversity data overview	• **Observation type:** Visual detections during 15 min survey under optimum weather conditions and during the species' main flight period. • **Response/Data type:** presence-absence data.
Type of predictors	Topopgraphic, land cover, vegetation (host species, cover, height, structure), management (disturbance), habitat type, landscape context
Conceptual model / Hypotheses	**Hypotheses about species-environment relationships based on previous studies:** Species are affected by resources (nectar plants), disturbances (land use and management), habitat type, terrain, vegetation structure, and landscape context.
Assumptions	We assumed that (i) relevant ecological drivers (or proxies) of species distributions are included, (ii) detectability does not change across habitat gradients, (iii) the species are at equilibrium with their environment, and (iv) sampling is adequate and representative.

SDM algorithms	• **Algorithms**: Logistic regression (generalised linear model GLM with binomial error distribution), hierarchical partitioning • **Model complexity**: We allowed quadratic relationships in GLMs, model complexity has been reduced by backward stepwise variable selection. Internal model validation by bootstrapping revealed only very slight overfitting. • **Ensembles**: N/A
Model workflow	• **Original study:** according to Hosmer and Lemeshow (2000): univariate analysis (identification of predictors with unimodal relationships, combination of categories if necessary). Likelihood-ratio-test to identify set of candidate predictors for multiple models. Selection of predictors in case of strong correlation (Spearman rho values of 0.5). Stepwise backward selection based on likelihood-ratio-test p-values (p_{in} = 0,05; p_{out} = 0.10). Model evaluation (AUC, R_{2N}, max_{kappa}); model validation by bootstrapping with 300 iterations. Response curves were produced for visualisation and understanding. • **Re-analysis study:** Stepwise backward selection based on AIC. Model evaluation (AUC with 95%-CI, R_{2N}, kappa, correct classification rate, sensitivity, specificity all for max_{kappa}); model validation by bootstrapping with 1000 iterations (AUC, R_{2N}, calibration plot), response curves, regression diagnostics (leverage plots, Cook's distance), check for spatial autocorrelation (Moran's I correlograms), variance partioning according to Borcard et al. (1992), MacNally & Walsh (2004) and Heikkinen et al. (2005).
Software	• **Software**: All analyses were conducetd using R (no version specified, R Core Team) with packages Hmisc (Harrell), lrm (i.e. predecessor of package rms, Harrell), hier.part (Walsh & MacNally), spdep (Bivand et al.), ncf (Bjornstad, 2016), sp (Pebesma et al.) (original study: SPSS 11.0$_{TM}$ and S-plus 6.1 with Design library (Harrell, 2001)(for validation)) • **Code availability**: code was not provided, but is available on demand from the authors • **Data availability**: data are not available
DATA	
Biodiversity data	• **Taxon names**: All (=2) species are given in both publications • **Taxonomic reference system**: We follow the taxonomy of the Red Data Books for Bavaria and Germany (e.g. Pretscher, 1998) • **Ecological level**: population level • **Data source:** Butterfly species were identified visually down to the species level by a trained surveyor. Plots were surveyed in two transect walks within a period of 15 min. The surveys were carried out under optimum weather conditions during the species main flight periods. • **Sampling design**: sample plots were selected following a random stratified scheme. Strata covered a range of six habitat types representing the gradient of habitats within the study area. • **Sample size**: The data set contained 118 sample plots (30x30 m_2) • **Absence data:** non-observation during two sampling periods in 2001 and 2002 was treated as absence.
Data partitioning	No data partitioning was applied, but internal validation by bootstrapping with 1000 replicates.

Predictor variables	• **Predictor variables**: o *Topography*: slope, sin and cos of exposition, potential solar radiation o *Land cover*: habitat type in six classes (fallow field/intensively managed grassland/extensively managed grassland/semi-arid grassland/fringe, thermophile/hedges and shrub) o *Disturbance*: Management type in six classes (mowing/cattle grazing and mowing/cattle grazing/shepherding/mulching/fallow), date of first annual management in five classes (before June 15/before July 15/before August 15/after August 15/fallow) o *Vegetation*: cover and height of vegetation layers (e.g. trees, shrubs, moss, herbs), presence/absence of specific nectar plants (among others: *Centaurea jacea, Scabiosa columbaria*), cover of bare ground o *Landscape context*: e.g. percentage of dry grasslands within a radius of 100 m, proportion of hedges within a radius of 25 m, cover of suitable habitat types within a radius of 25 m • **Data sources:** All predictor variables related to land use, vegetation and habitat type have been sampled during the field campaign, shortly after the butterflies have been sampled. Terrain variables have been calculated from a digital elevation model with 5 m resolution provided by the Bayerische Vermessungsverwaltung • **Spatial resolution and extent of raw data**: Data were collected and prepared at the same resolution as the biodiversity data (30x30 m$_2$-plots). Only habitat type and variables derived from the digital elevation model are available for the entire area; all other variables were only collected at the sample plots. • **Temporal resolution and extent of raw data**: Environmental data were collected directly after sampling butterflies on the plots. • **Data processing:** o *Topography*: slope, exposition and potential solar radiation were based on a 5 m digital elevation model. Calculations have been carried out with DiGeM 2.0 a program for digital terrain analysis (Conrad, 1998, 2002) and predecessor of SAGA-System for Automated Geoscientific Analyses (saga-gis.org, Conrad et al., 2015) o *Landscape context*: Variables were calculated based on GIS-analyses of the map of habitat types.
MODEL	
Variable pre-selection	Predictors were pre-selected based on their hypothesised ecological relevance for the distribution of these butterfly species.
Multicollinearity	We assessed collinearity using spearman rank correlation (Dormann et al., 2013). In case of Spearman rho values > 0.5 (Fielding& Haworth, 1995) only one predictor of the correlated ones was selected (i.e. the one with best performance in a univariate model).
Model settings	From the set of candidate predictors, we estimated univariate GLMs as well as GLMs with linear and quadratic terms (original study) according to (Hosmer & Lemeshow 2000).
Model selection	Backward selection based on AIC
Model estimates	We analysed model coefficients by comparing univariate and mutiple models, and we assessed the importance of predictors by hierarchical partitioning (Heikkinen et al.

	2005).
Non-independence	Spatial autocorrelation in model residuals was assessed using spline correlograms in the R package "ncf" (Bjornstad, 2016). (not original study, but re-analysis).
Threshold selection	Binary predictions for mapping and further calculations we used a threshold maximising kappa (Liu et al. 2005).
ASSESSMENT	
Performance statistics	Model evaluation (AUC with 95%-CI, R_{2N}, kappa, correct classification rate, sensitivity, specificity - all for $\text{max}_{\text{kappa}}$); Model validation by bootstrapping with 1000 iterations (AUC, R_{2N}, calibration plot), Additionally: regression diagnostics (leverage plots, Cook's distance), and check for spatial autocorrelation (Moran's I correlograms, cf. Dormann et al., 2007)
Plausibility checks	We checked model plausibility by assessing partial dependence plots.
PREDICTION	
Prediction output	Models based on area-wide available predictors were used to produce habitat suitability maps depicting occurrence probabilities. Binarised maps were used as the basis for graph-theory-based habitat connectivity analysis (Keitt et al. ,1997).
Uncertainty quantification	N/A

References

Conrad O. 1998. Die Ableitung hydrologisch relevanter Reliefparameter am Beispiel des Einzugsgebietes Linnengrund Diplomarbeit thesis, Universität Göttingen.
Conrad,O. 2002. DiGeM – Software for Digital Terrain Analysis. HP: http://www.geogr.uni-goettingen.de/pg/saga/digem/.
Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development 8, 1991-2007.
Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Daniel W. Kissling, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609-628.
Fielding AH, Haworth PF. 1995. Testing the generality of bird-habitat models. Conservation Biology 9, 1466-1481.
Harrell FE, Jr. 2001. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, New York.
Heikkinen RK, Luoto M, Kuussaari M, Pöyry J. 2005. New insights into butterfly-environment relationships using partitioning methods. Proceedings of the Royal Society of London / B 272, 2203-2210.
Hosmer DW, Lemeshow S. 2000. Applied logistic regression. Wiley, New York.
Keitt TH, Urban DL, Milne BT. 1997. Detecting critical scales in fragmented landscapes. Conservation Ecology 1, http://www.consecol.org/vol1/iss1/art4.

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

Mac Nally R, Walsh CJ. 2004. Hierarchical partitioning public-domain software. Biodiversity and Conservation 13, 659-660.

S9) Zurell et al. (2020): SDMs and JSDMs for Swiss forest birds and Swiss tree species

Zurell D, Zimmermann NE, Gross H, Baltensweiler A, Sattler T, Wüest R (2020) Testing species assemblage predictions from stacked and joint species distribution models. Journal of Biogeography 47: 101-113.

Background to case study:

This case study compared stacked species distribution models (S-SDMs) and joint species distribution models (JSDMs) in their ability to predict species assemblages. SDMs and JSDMs were used to predict the occurrence probabilities of different species. These predictions were then combined into species assemblage predictions using different approaches. Specifically, the study tested the effect of different stacking procedures (binary vs. probabilistic), the effect of applying macroecological constraints and ecological assembly rules, and the effect of bias correction in macroecological constraints on the accuracy of species assemblage predictions.

Below, we provide an example of how the Overview section could be formulated for the methods part of the main manuscript as well as the corresponding ODMAP table to be included in Online Supplementary Information.

Text example Overview section:

We describe the SDMs and JSDMs following the ODMAP (Overview, Data, Model, Assessment, Prediction) protocol for species distribution models. Here, we provide the **Overview** of the distribution models while the remaining ODMAP sections are detailed in Table S9.1 in the Supplementary Information.

The model objective for SDMs and JSDMs was to predict single species occurrence in space, both as continuous occurrence probabilities as well as binary maps of potential presence. For later analyses, these single species maps were stacked to predict species assemblages. SDMs and JSDMs were fitted for (i) forest bird species and (ii) bush and tree species in Switzerland. We restricted the models to the political boundaries of Switzerland (spatial extent: Longitude 5.76° E - 8.12° E, Latitude 45.70° N – 47.93° N). Bird presence-absence data at a 1x1 km spatial resolution were obtained from standardised monitoring carried out for the Swiss breeding bird atlas (Schmid et al., 1998), recorded over a four-year period (1993-1996). Tree species presence-absence data were obtained from standardised monitoring carried out for the Swiss National Forest Inventory (NFI) carried out between 2004-2006, and were aggregated to 100x100 m plot size to match the minimum grain of available environmental data.

We assumed that detection errors were negligible. Previous analyses of the Swiss breeding bird data have shown that the sampling approach ensures high species detectability of approximately 90% (Kéry & Schmid, 2006). In the forest inventory, tree individuals below a certain diameter at breast height are not recorded and we assume that this procedure does not bias the species identification.

Based on previous studies, we selected climate, topography and vegetation structure as important environmental predictor variables for bird and tree species in Switzerland. SDMs and JSDMs were estimated using presence/absence data as response variable to predict species-specific occurrence probabilities per site. Additionally, we estimated SDMs using species richness as response variable to predict species richness per site, which then served as macroecological constraint for the application of ecological assembly rules. SDMs and richness models were fitted using four different algorithms: generalised linear models (GLM), generalised additive models (GAM), boosted regression trees (BRT), and random forests (RF) with a binomial error distribution (with logit link) for SDMs and a Poisson error distribution (with log link) for richness models. JSDMs were fitted with a latent variable model and a binomial error distribution (with probit link). Model settings were chosen to yield intermediately complex response surfaces.

Prior to model building, all predictor variables were standardised. In each model, we only included the five most important and weakly correlated variables. Univariate variable importance for each predictor was assessed in a 5-fold spatial block cross-validation design. Ensemble predictions from SDMs and richness models were derived using un-weighted ensemble means. Predictive model performance was assessed using a 5-fold spatial block cross-validation.

All analyses were conducted using R version 3.3.2 (R Core Team, 2016) with packages sperrorest (Brenning, 2012), mgcv (Wood, 2011), gbm (Ridgeway, 2013), dismo (Hijmans et al., 2017), randomForest (Liaw & Wiener, 2002), boral (Hui, 2016), ncf (Bjornstad, 2016), PresenceAbsence (Freeman & Moisen, 2008), ecospat (Broennimann et al., 2016) and lme4 (Bates et al., 2015). All codes are provided in the Online Supplementary Information and on github (https://github.com/damariszurell/SSDM-JSDM). Data are available from Dryad (https://doi.org/10.5061/dryad.k88v330).

Supplementary Table S9.1. ODMAP protocol. Details on Data, Model, Assessment, Prediction. For Overview section, please refer to main text.

ODMAP element	**Contents**
OVERVIEW	
Authorship	• **Authors:** Damaris Zurell, Niklaus E. Zimmermann, Helge Gross, Andri Baltensweiler, Thomas Sattler, Rafael O. Wüest • **Contact email:** damaris@zurell.de • **Title:** Testing species assemblage predictions from stacked and joint species distribution models • **DOI:** 10.1111/jbi.13608
Model objective	• **Objective**: Mapping/interpolation • **Target outputs**: continuous occurrence probabilities and binary maps of potential presence
Taxon	• (i) forest bird species, and (ii) bush and tree species

Location	Switzerland
Scale of analysis	- **Spatial extent (Lon/Lat):** Longitude 5.76° E - 8.12° E, Latitude 45.70° N – 47.93° N - **Spatial resolution:** (i) 1 km x1 km for forest bird species, and (ii) 100 m x 100 m for bush and tree species - **Temporal resolution and extent**: We just modelled a single time slice for birds (1993-1996) and for trees (2004-2006). - **Type of extent boundary:** political (Switzerland)
Biodiversity data overview	- **Observation type**: standardised monitoring - **Response/Data type**: presence/absence data
Type of predictors	- Climatic, topographic, vegetation structure
Conceptual model / Hypotheses	- Based on previous studies, we tested climate, topography and vegetation structure as important environmental predictor variables for bird and tree species in Switzerland in an exploratory way.
Assumptions	We assumed that species are at pseudo-equilibrium with the environment. Also, we assumed that detection errors were negligible. Previous analyses of the Swiss breeding bird data have shown that the sampling approach ensures high species detectability of approximately 90% (Kéry & Schmid, 2006). In the forest inventory, tree individuals below a certain diameter at breast height are not recorded and we assume that this procedure does not bias the species identification.
SDM algorithms	- **Algorithms**: SDMs and richness models were fitted using four different algorithms: generalised linear models (GLM), generalised additive models (GAM), boosted regression trees (BRT), and random forests (RF) with a binomial error distribution (with logit link) for SDMs and a Poisson error distribution (with log link) for richness models. JSDMs were fitted with a latent variable model and a binomial error distribution (with probit link). - **Model complexity**: Model settings were chosen to yield intermediately complex response surfaces. We allowed quadratic relationships in GLMs and JSDMs, and restricted GAMs, BRTs and RFs such that these would not overfit too much. - **Ensembles**: We combined the four SDMs to ensemble SDM predictions.
Model workflow	Prior to model building, all predictor variables were standardised. In each model, we only included the five most important and weakly correlated variables. Univariate variable importance for each predictor was assessed in a 5-fold spatial block cross-validation design. Ensemble predictions from SDMs and richness models were derived using un-weighted ensemble means. Predictive model performance was assessed using a 5-fold spatial block cross-validation.
Software	- **Software**: All analyses were conducted using R version 3.3.2 (R Core Team, 2016) with packages sperrorest (Brenning, 2012), mgcv (Wood, 2011), gbm (Ridgeway, 2013), dismo (Hijmans et al., 2017), randomForest (Liaw & Wiener, 2002), boral (Hui, 2016), ncf (Bjornstad, 2016), PresenceAbsence (Freeman & Moisen, 2008), ecospat (Broennimann et al., 2016) and lme4 (Bates et al., 2015). - **Code availability:** All codes were provided in the Online Supplementary Information and on github (https://github.com/damariszurell/SSDM-JSDM). - **Data availability**: Data are available from Dryad

	(https://doi.org/10.5061/dryad.k88v330).
DATA	
Biodiversity data	• **Taxon names:** All species are listed in Table S1 and S2 of the original publication. • **Taxonomic reference system:** We follow the taxonomy of the Swiss breeding bird atlas and the Swiss National Forest Inventory. • **Ecological level:** population level • **Data source**: Bird presence-absence data at a 1x1 km spatial resolution were obtained from the Swiss breeding bird atlas (Schmid et al., 1998). Tree species presence-absence data were obtained from the Swiss National Forest Inventory (NFI). • **Sampling design**: Bird atlas data were recorded over a four-year period (1993-1996) in usually three visits per year (two above the treeline) using a simplified territory mapping approach. The NFI samples Switzerland on a regular grid (spacing 1.4 km), and in case the sample falls into forest it records forest characteristics in a maximal area of 50x50 m (Brassel & Lischke, 2001). • **Sample size**: The bird data set contained 2535 1x1 km cells with a total number of 56 forest bird species and prevalence ranging 0.03-0.73. The tree data set contained 6946 100x100 m cells with a total number of 63 tree and shrub species and prevalence ranging 0.01-0.79. • **Country mask:** we clipped all data to the political boundary of Switzerland. • **Scaling**: We aggregated the NFI presence-absence data to 100x100 m plot size to match the minimum spatial grain of available environmental data. • **Data filtering**: We only considered species with at least 50 presences. • **Absence data**: The Swiss breeding bird atlas and the NFI contains presence and absence data. In the Swiss breeding bird data, species are listed as absent in a site if they were not encountered within the 2-3 visits per breeding season in the four successive years of recording. Previous analyses have shown that this sampling approach ensures high species detectability of approximately 90% (Kéry & Schmid, 2006). In the NFI plots, all tree and bush individuals above a certain diameter at breast height are recorded. We have no information about potential species biases in unrecorded small trees and bushes.
Data partitioning	We randomly selected 70% of data (1774 cells for birds and 4862 for tree species) for model building and 30% (761 cells for birds and 2084 cells for tree species) for validation of the community predictions. Predictive model performance for single species was assessed using a 5-fold spatial block cross-validation design. Therefore, for each dataset we split the study region into five rectangular tiles (R package "sperrorest"). The resulting sample sizes per tile ranged 222-568 cells for forest birds and 570-1440 cells for tree species.
Environmental data/predictor variables	• **Predictor variables:** ○ *Topography*: slope, aspect, topographic position index (TPI), topographic wetness index (TWI), potential monthly solar radiation ○ *Climate*: 19 bioclimatic predictors, degree days, potential evapotranspiration (PET), moisture balance (MBAL), moisture index (MIND) ○ *Vertical vegetation structure*: LiDAR-derived average height, height standard deviation, height coefficient of variation, as well as 10th, 25th

    and 95th height percentiles, canopy cover (COV), canopy density (DNS), foliage height diversity (FHD), understory height diversity (UHD)
  - *Horizontal vegetation structure (for bird data only)*: edge length between two height classes, clumpiness for different height classes
- **Data sources**: We prepared the environmental predictor variables climate, topography and LiDAR-derived vegetation structure. Monthly average climate data were obtained from the Federal Office of Meteorology and Climatology MeteoSwiss (www.meteosuisse.ch). Topography and LiDAR data were obtained from the Swiss Federal Office of Topography.
- **Spatial resolution and extent of raw data**: The raw resolution of topography and climate was 100 m. The raw LIDAR point clouds had a nominal footprint of 0.3 m. Data were prepared at the same resolution as the species data, meaning at 1x1 km for analyses of forest birds and at 100x100 m for analyses of tree species.
- **Temporal resolution and extent of raw data:** Climate data stem from the period 1981-1990. LiDAR raw data were measured in the years 2000–2007.
- **Geographic projection:** Swiss coordinate system (Swiss grid)
- **Data processing**:
  - *Topography*: Slope, aspect, TPI and TWI (Wilson & Gallant, 2000) were based on a 100 m digital elevation model. TPI in a cell corresponds to the difference of the focal cell to the mean of its eight surrounding cells, with negative values indicating a depression, positive values a rise. Potential monthly solar radiation was calculated following Hofierka et al. (2002).
  - *Climate*: Monthly average climate data for the period 1981-1990 were generated by interpolating ca. 300 MeteoSwiss station data to a resolution of 100x100m using the Daymet software (Thornton et al., 1997). From these, we derived 19 bioclimatic predictors (http://worldclim.org/bioclim) as well as so-called degree days. Degree days constitute the sum of all monthly temperature values greater than a given threshold temperature multiplied by the total number of days (with thresholds 0°C and 5°C for DDEG0 and DDEG5). We calculated PET using radiation as proposed by Makkink (1957), as this method was shown to best approximate PET in Switzerland (Xu & Singh, 2002). For precipitation and PET we calculated summer (April to September) and winter (October to March) averages as well as their ratio. We calculated the MBAL as the difference between precipitation and PET, and MIND as the ratio between PET and precipitation. For the analyses of forest bird species, the topographic and climatic data were aggregated to 1x1 km grids using the mean as aggregate-function such that the grid cell-size matched the bird survey data.
  - *Vertical vegetation structure*: LiDAR-derived data was processed with the LAStools software (Isenburg, 2015). The Swiss-wide LiDAR dataset consists of discrete first and last pulse returns with a nominal footprint of 0.3 m and a point density of 0.5 points/m$_2$ (Artuso et al., 2003). LiDAR variables were generated separately as raster datasets for both the 100x100 m grid that matches the tree species NFI data, as well as the 1x1 km grid that matches the bird survey data. From the terrain corrected and classified LiDAR point cloud (heights of classified vegetation LiDAR returns minus interpolated DTM heights), we calculated the average height, their standard deviation and coefficient of variation as well as the corresponding 10th, 25th and 95th percentiles per 100 m and 1 km pixel, respectively. In addition, we also

	derived the canopy cover (percentage of first returns above 1m; COV) and canopy density (ratio of all returns above 1m divided by all returns; DNS). We further calculated both standard deviation and coefficient of variation in order to characterize vertical variation in LiDAR returns. In addition, we estimated the so-called foliage height diversity (FHD), which is the Shannon diversity index based on 5m vertical bins as $H = p_i \ln(p_i)$, where p_i is the proportion of LiDAR returns in the 5m bin i. Additionally, we calculated understory height diversity (UHD) for the 1x1 km grid as we considered these important predictors for bird distributions in Swiss forests. UHD was derived analogously to FHD but using 1m bins restricted to below 12m (Zellweger et al., 2016) as well as the ratio between the 95th and 25th percentile. o *Horizontal vegetation structure*: For the bird data at 1x1 km resolution, we derived LiDAR variables accounting for edge effects and fragmentation in order to describe the horizontal structural heterogeneity of the vegetation (Zellweger et al., 2013). Based on the terrain corrected LiDAR point cloud we generated a gridded Canopy Height Model (CHM) with a grid size of 20 m. The CHM was classified into three classes, which are non-forest (vegetation height < 1m), understory/midstory (1 – 12m) and canopy (> 12m). The length of edges between two height classes (e.g. non-forest/canopy) was calculated for each grid cell and summed up for the 1x1 km grid of the bird survey data. To measure the spatial dispersion or aggregation of the vegetation height classes understory/midstory and canopy, a clumpiness index was calculated for the 1x1 km grid using the software FRAGSTATS (McGarigal et al., 2012). o Because the forest bird data were recorded 1993-1996 and the LiDAR data were recorded after 2000, there is a temporal mismatch between species data and vegetation data. Generally, the forest laws in Switzerland are very strict and we can, thus, rule out any major changes in the vegetation structure between these two periods. The main exception is storm damage due to the cyclone "Lothar" in 1999. In storm-damaged sites we can expect differences in vegetation structure between the recording periods of the bird survey data (before the cyclone) and the LiDAR data (after the cyclone). Hence, we removed all storm-damaged sites (n=10) from the analyses to avoid mismatches in vegetation structure.
MODEL	
Variable pre-selection	In each model, we only included the five most important and weakly correlated variables obtained from the cross-validated univariate variable importance.
Multicollinearity	We reduced the predictor set to variables with bivariate Spearman correlations $\|r\|<0.7$, retaining those variables from highly correlated pairs with higher cross-validated univariate importance (Dormann et al., 2013). Univariate variable importance for each predictor was assessed in a 5-fold spatial block cross-validation design by estimating univariate GAMs on 4 of 5 folds (with logit link for species occurrences and log link for species richness) and cross-predicting these to the left-out fold with 5 repetitions. From the cross-predictions, we calculated the percentage of explained deviance. We thus obtained a predictor ranking for each single species and removed the less important variables from pairs of highly correlated variables. However, JSDMs require a global set of predictor variables

	and we thus selected those five variables with highest mean cross-validated univariate importance among all species. SDMs and richness models were run with the same set of global predictors as in JSDMs. In a sensitivity analysis we additionally estimated SDMs and richness models using the five most important variables selected individually for each species and species richness, respectively. This did not change the overall results.
Model settings	• Model settings SDMs and richness models: GLMs were fitted with linear and quadratic terms and GAMs were fitted with nonparametric cubic smoothing splines with up to four degrees of freedom. BRTs were estimated with a tree complexity of 2, a bag fraction of 0.75 and a variable learning rate such that 1000-5000 trees were fitted (Elith et al., 2008). Random forests were fitted with 1000 trees, and a minimum nodesize of 20. • Model settings JSDMs: JSDMs were fitted with a latent variable model and a binomial error distribution (with probit link) (with R package "boral"; Hui, 2016). In all JSDMs we included linear and quadratic terms and five latent variables. For forest bird species, JSDMs were run with 50000 iterations, a burnin of 20000 and a thinning rate of 50. For tree species, JSDMs were run with 100000 iterations, a burnin of 50000 and a thinning rate of 50. Convergence was assessed using the Geweke convergence diagnostic.
Model estimates	We did not analyse model coefficients in depth, but compared for each species how often each predictor was selected among the five most important variables.
Model averaging / Ensembles	Consensus predictions from SDMs and richness models, respectively, were generated using un-weighted ensemble means.
Non-independence	Spatial autocorrelation in model residuals was assessed using spline correlograms in the R package "ncf" (Bjornstad, 2016).
Threshold selection	Binary predictions were derived by using the TSS (true skill statistic)-maximisation threshold. Maximising TSS is equivalent to maximising the sum of sensitivity and specificity.
ASSESSMENT	
Performance statistics	Predictive model performance on validation data (based on the 5-fold spatial block cross-validation) was assessed using four different performance measures: area under the receiver operating characteristic curve (AUC), true skill statistic (TSS), sensitivity and specificity. The latter three measures constitute threshold-dependent performance measures and we calculated them using a TSS-maximisation threshold.
Plausibility checks	In our pre-analyses, we used inflated response curves to understand model behaviour for different parameter settings, and based on these checks decided for intermediate model complexity.
PREDICTION	
Prediction output	**Prediction unit**: For further analyses, we used continuous predictions of occurrence probability per species and site as well as predicted presence per species and site that was obtained by binarising the predicted occurrence probabilities using the TSS-maximisation threshold.

Uncertainty quantification	In SDMs, we accounted for algorithmic uncertainty by applying an ensemble approach averaging over four different SDM algorithms. In a sensitivity analysis, we also compared predictions from GLMs only against JSDMs.