SUPPORTING INFORMATION

Underprediction of extirpation and colonisation following climate and land-use change using species distribution models

Alistair G. Auffret, Hedvig Nenzèn, Ester Polaina

Contains model outputs for analyses used in the paper, as well as visualisations relating to niche overlap and extrapolation to new environments, and an ODMAP protocol. The published data files and R code can be used to produce these tables and figures. See main manuscript and dataset for details regarding the variables used.

SUPPLEMENTARY TABLES S1-S8

Table S1. Output of linear model assessing relationship between TSS scores of species distribution models of 84 plant species in four Swedish provinces, assessed using traditional cross-validation and independent temporal validation. Model R^2 = 0.11.

Predictor	Estimate	Standard Error	T value	P value
(Intercept)	-0.2044125	0.15316312	-1.3346067	0.18569914
TSS_cross	0.74618637	0.23385856	3.19075925	0.00201147

Table S2. Outputs of quasi-binomial generalised linear models assessing relationships between fraction of 84 species' predicted colonisations, persistences and extirpations by SDMs that were observed following changes in climate and land use. Separate models for each response were created, swapping species temperature range index (STRI) and species mean temperature index (SMTI).

Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation				
STRI - R^2 = 0.21				
(Intercept)	7.25258282	1.64480184	4.4093961	3.20E-05
grass.spec	-0.0703093	0.06310921	-1.11409	0.26857657
forest.spec	-0.148624	0.06020398	-2.4686733	0.01569255
SCI_TempRange	-0.2004026	0.04885898	-4.1016529	9.80E-05
SMTI - R^2 = 0.20				
(Intercept)	-5.7953201	1.59238052	-3.6394066	0.00048275
grass.spec	0.04749722	0.06950586	0.68335558	0.4963566
forest.spec	-0.0662725	0.05873097	-1.1284084	0.26252014
SCI_TempMean	0.98556662	0.24085009	4.09203347	0.00010139
Persistence				
STRI - R^2 = 0.35				
(Intercept)	10.5492877	1.914638	5.50980794	4.24E-07
grass.spec	-0.0347192	0.06408142	-0.5417984	0.58946328
forest.spec	-0.1728837	0.05429661	-3.1840613	0.00206983
SCI_TempRange	-0.2750511	0.05505493	-4.9959396	3.37E-06
SMTI - R^2 = 0.35				
(Intercept)	-6.6563999	1.46530658	-4.5426671	1.94E-05
grass.spec	0.11927748	0.07269565	1.64077878	0.10476951
forest.spec	-0.0671974	0.05369145	-1.2515474	0.21438209
SCI_TempMean	1.24487881	0.22862657	5.44503128	5.54E-07
Extirpation				
STRI - R^2 = 0.26				
(Intercept)	-7.6948892	1.42137119	-5.4137085	6.29E-07
grass.spec	0.07010773	0.04787965	1.46424888	0.1470446
forest.spec	0.15916228	0.06136845	2.59355234	0.01129265
SCI_TempRange	0.19624405	0.04259687	4.60700669	1.52E-05
SMTI - R^2 = 0.17				
(Intercept)	2.67376859	1.21392994	2.20257241	0.030503
grass.spec	-0.0032198	0.0540946	-0.0595216	0.95268507
forest.spec	0.1063345	0.06413515	1.65797529	0.10123885
SCI_TempMean	-0.6244112	0.18141675	-3.4418606	0.00092138

Table S3. Outputs of quasi-binomial generalised linear models assessing relationships between fraction of 84 species' predicted colonisations, persistences and extirpations by SDMs that were observed following changes in climate and land use, including values of Schoener's D niche overlap. Separate models for each response were created, swapping species temperature range index (STRI) and species mean temperature index (SMTI).				
Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation				
STRI - R^2 = 0.21				
(Intercept)	7.72023852	1.82262198	4.23578701	6.11E-05
grass.spec	-0.0655357	0.06430129	-1.019197	0.3112206
forest.spec	-0.1428518	0.06083096	-2.3483397	0.02135862
SCI_TempRange	-0.1832253	0.05458131	-3.3569243	0.00121422
niche.overlap	-1.8339065	2.69446692	-0.6806194	0.49810234
SMTI - R^2 = 0.21				
(Intercept)	-4.2166809	2.79513396	-1.5085792	0.13539542
grass.spec	0.04409094	0.07033403	0.62687918	0.53254592
forest.spec	-0.0674383	0.05917235	-1.1396921	0.25785837
SCI_TempMean	0.90379082	0.26853434	3.36564334	0.00118118
niche.overlap	-1.8503819	2.74269413	-0.6746585	0.5018624
Persistence				
STRI - R^2 = 0.35				
(Intercept)	10.8632493	2.17279496	4.99966608	3.38E-06
grass.spec	-0.033398	0.06502149	-0.513645	0.60893383
forest.spec	-0.1715575	0.05493121	-3.1231341	0.00250117
SCI_TempRange	-0.2676363	0.05969643	-4.4832877	2.45E-05
niche.overlap	-0.9671108	2.92846108	-0.3302454	0.74208873
SMTI - R^2 = 0.36				
(Intercept)	-4.9732131	2.5857073	-1.9233473	0.05803931
grass.spec	0.11543298	0.07430847	1.55342967	0.12431772
forest.spec	-0.0682343	0.05495169	-1.2417145	0.2180159
SCI_TempMean	1.19405522	0.24247097	4.92452856	4.54E-06
niche.overlap	-2.3286516	2.98139869	-0.7810601	0.43710148
Extirpation				
STRI - R^2 = 0.28				
(Intercept)	-8.5632259	1.67125092	-5.1238421	2.07E-06
grass.spec	0.07273768	0.04841938	1.50224315	0.13702119
forest.spec	0.14521535	0.0621166	2.33778665	0.02192988
SCI_TempRange	0.17086158	0.04744334	3.60138206	0.0005511
niche.overlap	2.97780348	2.63738081	1.12907604	0.26228265
SMTI - R^2 = 0.20				
(Intercept)	-1.0587583	2.37778205	-0.4452714	0.65734141
grass.spec	0.01951427	0.05539486	0.35227581	0.72556947
forest.spec	0.09435129	0.06372759	1.48054061	0.14270634
SCI_TempMean	-0.4684798	0.19936648	-2.3498422	0.02127837
niche.overlap	4.80093738	2.70812727	1.77278868	0.08011848

Table S4. Outputs of quasi-binomial generalised linear models assessing relationships between fraction of observed colonisations (n=1032 grid squares), persistences (n=1022 grid squares) and extirpations (n=969 grid squares) that were correctly predicted by SDMs of 84 species, with the level of grassland abandonment and temperature change that have occurred within a grid square. Models also included control variables for microclimatic variability, latitude, and spatial autocorrelation.				
Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation – R^2 = 0.05				
(Intercept)	0.86661945	0.91681272	0.94525243	0.34475278
temp_change	-0.2714579	0.65948496	-0.411621	0.68070326
micro_sd	-0.0016604	0.00991366	-0.1674832	0.86702286
grass_aba	-0.4229192	0.34415929	-1.2288473	0.21941112
lat.res	-4.24E-06	1.03E-06	-4.1372446	3.80E-05
pcnm1	4.34445942	2.71991114	1.59727991	0.11051155
pcnm2	-5.9220564	2.66649793	-2.2209117	0.02657488
Persistence – R^2 = 0.19				
(Intercept)	6.56363885	0.96080058	6.83142679	1.45E-11
temp_change	-3.6406754	0.68771647	-5.293861	1.47E-07
micro_sd	-0.0190851	0.00995799	-1.9165641	0.05557391
grass_aba	-1.2803371	0.3297562	-3.882678	0.00010999
lat.res	-1.06E-05	1.03E-06	-10.205504	2.40E-23
pcnm1	13.1023492	2.93208769	4.46860756	8.76E-06
pcnm2	-21.40872	2.70822481	-7.9050749	6.95E-15
Extirpation– R^2 = 0.04				
(Intercept)	-1.7252947	1.2310828	-1.4014449	0.1614036
temp_change	0.46640475	0.87782946	0.53131591	0.59532254
micro_sd	0.00046068	0.01204248	0.03825436	0.96949282
grass_aba	0.12395295	0.44578301	0.27805669	0.78102859
lat.res	4.73E-06	1.31E-06	3.59421454	0.00034188
pcnm1	-1.6527775	3.79516254	-0.4354958	0.66330025
pcnm2	6.80704822	3.41383796	1.99395762	0.04643862

Table S5. Sensitivity analyses – species level (lower threshold for predicted presence). Outputs of quasi-binomial generalised linear models assessing relationships between fraction of 84 species' predicted colonisations, persistences and extirpations by SDMs that were observed following changes in climate and land use. Separate models for each response were created, swapping species temperature range index (STRI) and species mean temperature index (SMTI).

Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation				
STRI - R^2 = 0.22				
(Intercept)	7.95225912	1.74333018	4.56153355	1.81E-05
grass.spec	-0.0655155	0.06404185	-1.0230104	0.3093865
forest.spec	-0.1500961	0.06054765	-2.4789747	0.01527859
SCI_TempRange	-0.2151139	0.05149526	-4.1773544	7.47E-05
SMTI - R^2 = 0.21				
(Intercept)	-5.7200547	1.60826624	-3.5566591	0.00063456
grass.spec	0.05981987	0.07181571	0.83296357	0.40734519
forest.spec	-0.0632182	0.05919952	-1.0678831	0.2887851
SCI_TempMean	1.00520335	0.24403136	4.1191564	9.20E-05
Persistence				
STRI - R^2 = 0.36				
(Intercept)	11.2138016	2.01975824	5.55205142	3.57E-07
grass.spec	-0.0236162	0.06570889	-0.359406	0.72023832
forest.spec	-0.1696044	0.05370587	-3.1580242	0.00224112
SCI_TempRange	-0.2878421	0.05789941	-4.9714173	3.72E-06
SMTI - R^2 = 0.38				
(Intercept)	-6.7759979	1.45301887	-4.6633929	1.23E-05
grass.spec	0.14422282	0.07512162	1.91985764	0.05844049
forest.spec	-0.0599587	0.05322049	-1.1266092	0.26327584
SCI_TempMean	1.29923898	0.22788365	5.70132596	1.92E-07
Extirpation				
STRI - R^2 = 0.27				
(Intercept)	-8.2411874	1.53646765	-5.3637233	7.72E-07
grass.spec	0.05485525	0.05020437	1.09263897	0.27783179
forest.spec	0.16545113	0.06231833	2.65493534	0.00956868
SCI_TempRange	0.20518238	0.04575507	4.48436392	2.42E-05
SMTI - R^2 = 0.18				
(Intercept)	2.53927049	1.26423349	2.0085455	0.04795936
grass.spec	-0.021547	0.05801186	-0.3714234	0.7113041
forest.spec	0.1104268	0.06530366	1.69097404	0.09473441
SCI_TempMean	-0.6405986	0.18954164	-3.3797247	0.00112389

Table S6. Sensitivity analyses – species level (higher threshold for predicted presence). Outputs of quasi-binomial generalised linear models assessing relationships between fraction of 84 species' predicted colonisations, persistences and extirpations by SDMs that were observed following changes in climate and land use. Separate models for each response were created, swapping species temperature range index (STRI) and species mean temperature index (SMTI).

Predictor	Estimate	Standard Error	T value	P value
Colonisation				
STRI - R^2 = 0.14				
(Intercept)	5.11907593	1.46017499	3.50579621	0.00074927
grass.spec	-0.0545538	0.06133717	-0.8894082	0.37645047
forest.spec	-0.1240816	0.05932065	-2.0917095	0.03963655
SCI_TempRange	-0.1488518	0.04400738	-3.3824289	0.00111427
SMTI - R^2 = 0.16				
(Intercept)	-5.3452039	1.54191193	-3.4666078	0.00085075
grass.spec	0.03878733	0.06564365	0.5908772	0.55626843
forest.spec	-0.0597143	0.05832528	-1.0238144	0.30900888
SCI_TempMean	0.85177084	0.23129417	3.68262989	0.00041786
Persistence				
STRI - R^2 = 0.24				
(Intercept)	8.02115985	1.81708068	4.41431134	3.14E-05
grass.spec	-0.0284545	0.06843404	-0.4157947	0.67867346
forest.spec	-0.1358071	0.06022024	-2.2551735	0.02685471
SCI_TempRange	-0.2150862	0.05294257	-4.0626321	0.00011256
SMTI - R^2 = 0.25				
(Intercept)	-6.1012186	1.66998712	-3.6534525	0.00046068
grass.spec	0.09185986	0.07786897	1.1796722	0.24162679
forest.spec	-0.0499514	0.06213305	-0.8039423	0.42381278
SCI_TempMean	1.08276671	0.25784344	4.19931845	6.90E-05
Extirpation				
STRI - R^2 = 0.22				
(Intercept)	-6.7217223	1.39868528	-4.8057432	7.10E-06
grass.spec	0.05255522	0.04998513	1.05141708	0.2962327
forest.spec	0.1245168	0.06583462	1.89135744	0.06219597
SCI_TempRange	0.18456637	0.04266963	4.32547411	4.36E-05
SMTI - R^2 = 0.14				
(Intercept)	3.19740056	1.26017131	2.53727452	0.01311514
grass.spec	-0.017286	0.05381806	-0.3211938	0.74890114
forest.spec	0.07459381	0.06672534	1.11792328	0.26694565
SCI_TempMean	-0.6179801	0.18687737	-3.3068749	0.0014146

Table S7. Sensitivity analyses – grid-cell level (lower threshold for predicted presence). Outputs of quasi-binomial generalised linear models assessing relationships between fraction of observed colonisations (n=1032 grid squares), persistences (n=1022 grid squares) and extirpations (n=969 grid squares) that were correctly predicted by SDMs of 84 species, with the level of grassland abandonment and temperature change that have occurred within a grid square. Models also included control variables for microclimatic variability, latitude, and spatial autocorrelation.

Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation – R^2 = 0.05				
(Intercept)	0.25839923	0.94471264	0.27352151	0.78450746
temp_change	0.28729102	0.67982156	0.42259769	0.67267745
micro_sd	-0.0001871	0.01026186	-0.0182374	0.985453
grass_aba	-0.3757526	0.35734957	-1.0514987	0.29327742
lat.res	-3.56E-06	1.06E-06	-3.3498376	0.00083806
pcnm1	4.16433034	2.80937414	1.48229824	0.13856832
pcnm2	-4.0831456	2.75833773	-1.4802921	0.1391026
Persistence – R^2 = 0.19				
(Intercept)	6.6721244	1.00858585	6.61532618	5.99E-11
temp_change	-3.5541598	0.7218549	-4.9236484	9.91E-07
micro_sd	-0.019358	0.01046799	-1.8492585	0.06471109
grass_aba	-1.2537368	0.34771912	-3.6056022	0.00032658
lat.res	-1.11E-05	1.09E-06	-10.171186	3.31E-23
pcnm1	14.3009068	3.07849072	4.64542793	3.84E-06
pcnm2	-22.43487	2.84183719	-7.8944952	7.53E-15
Extirpation– R^2 = 0.06				
(Intercept)	-2.4876777	1.30408878	-1.9075984	0.05674068
temp_change	0.74630947	0.92987577	0.80259051	0.42240969
micro_sd	0.00994658	0.01273681	0.78093156	0.43503487
grass_aba	0.05841335	0.47959179	0.12179806	0.90308438
lat.res	5.92E-06	1.40E-06	4.22072384	2.67E-05
pcnm1	-4.3759688	4.01475314	-1.0899721	0.2759984
pcnm2	8.91064016	3.6212099	2.46068038	0.01404214

Table S8. Sensitivity analyses – grid-cell level (higher threshold for predicted presence). Outputs of quasi-binomial generalised linear models assessing relationships between fraction of observed colonisations (n=1032 grid squares), persistences (n=1022 grid squares) and extirpations (n=969 grid squares) that were correctly predicted by SDMs of 84 species, with the level of grassland abandonment and temperature change that have occurred within a grid square. Models also included control variables for microclimatic variability, latitude, and spatial autocorrelation.

Predictor	**Estimate**	**Standard Error**	**T value**	**P value**
Colonisation – R^2 = 0.04				
(Intercept)	0.65718984	0.91330927	0.71956988	0.47195389
temp_change	-0.4416561	0.65657845	-0.6726631	0.50131325
micro_sd	0.00326376	0.00987604	0.33047285	0.74111024
grass_aba	-0.128509	0.34321732	-0.3744245	0.70816598
lat.res	-4.21E-06	1.01E-06	-4.1617203	3.42E-05
pcnm1	4.74171706	2.70460911	1.75319866	0.07986682
pcnm2	-6.8898002	2.64481884	-2.6050178	0.00932
Persistence – R^2 = 0.16				
(Intercept)	6.27210392	0.87283797	7.18587427	1.29E-12
temp_change	-3.8379024	0.62472043	-6.1433919	1.16E-09
micro_sd	-0.0031671	0.00911739	-0.3473679	0.72838695
grass_aba	-1.344419	0.30211121	-4.4500798	9.53E-06
lat.res	-9.59E-06	9.35E-07	-10.251637	1.56E-23
pcnm1	13.9748083	2.66672894	5.24043074	1.95E-07
pcnm2	-21.053817	2.4665577	-8.5357085	5.02E-17
Extirpation– R^2 = 0.04				
(Intercept)	-0.840486	1.12462722	-0.7473463	0.4550372
temp_change	0.17158168	0.80163386	0.21403996	0.83056125
micro_sd	-0.0070115	0.01108427	-0.6325638	0.52716886
grass_aba	0.61431212	0.40513612	1.51631043	0.12976939
lat.res	4.05E-06	1.19E-06	3.38529665	0.00073976
pcnm1	-0.6052227	3.47156252	-0.1743373	0.86163709
pcnm2	7.5498085	3.10574298	2.43091864	0.01524242

SUPPLEMENTARY FIGURES S1-S2

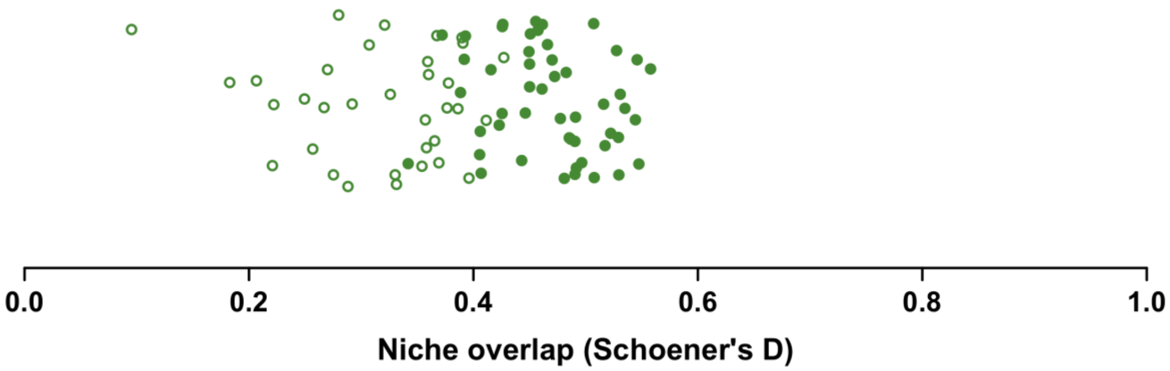

Figure S1. Values of Schoener's D, a measure of niche overlap in 84 plant species across three provinces of Sweden. Niches were calculated using values of land cover (proportion of arable, open, forest and water), soil (proportion of peat, sand-gravel, rock, silt/clay, till other) and climate (mean of monthly mean, minimum and maximum temperature and precipitation) in the historical and modern study periods. Possible values range between 0 (complete niche separation) and 1 (complete niche overlap). Filled circles indicate overlap significant at the P<0.05 level. Random jitter was added to the Y-axis to facilitate visualisation.

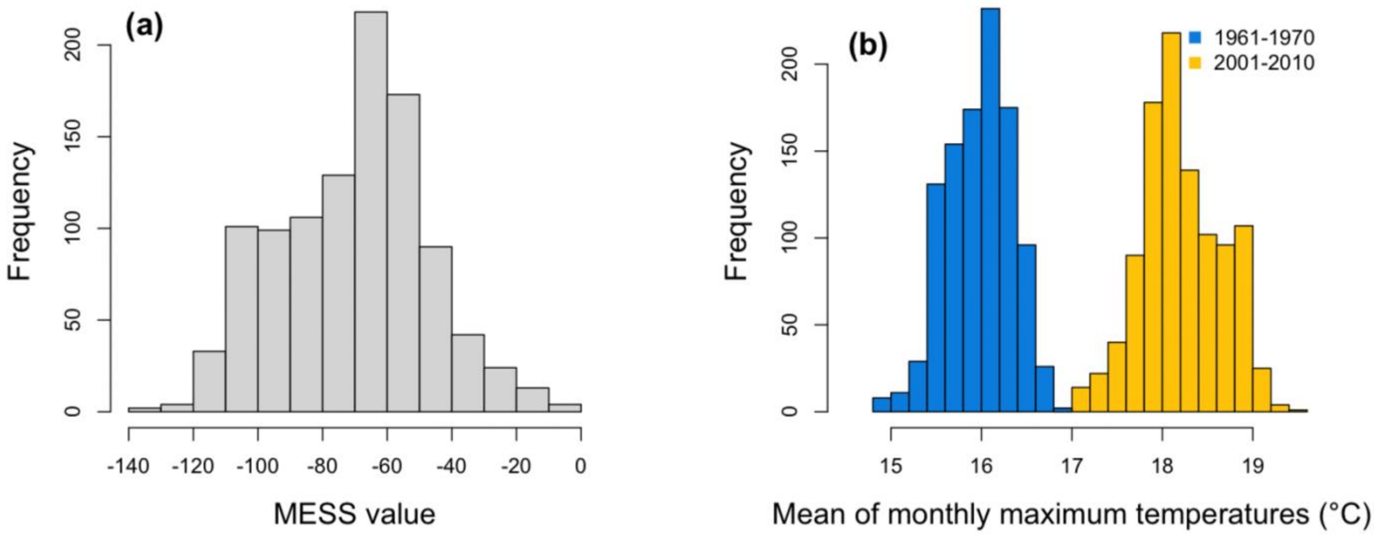

Figure S2. (a) Histogram showing number of grid squares of different MESS (multivariate similarity surface) values in three provinces across Sweden based on land use proportion of arable, open and forest) and climate (mean of maximum temperature and precipitation) in the historical and modern period. All values are negative, indicating extrapolation across the whole study area. (b) Histograms visualising the separation of mean monthly maximum temperatures between the historical and modern period. These differences were the sole reason for negative MESS values in 1015 of 1038 grid cells, while the lack of any open land in some grid cells in the historical period contributed to extrapolation in the remaining 23 cells.

Supplementary Text

Underprediction of extirpation and colonisation following climate and land-use change using species distribution models

– ODMAP Protocol –

Alistair G. Auffret, Hedvig Nenzén, Ester Polaina

2024-03-06

Overview

Authorship

Contact : alistair.auffret@slu.se

Study link: Accepted in Diversity and Distributions

Model objective

Model objective: Forecast and transfer

Target output: Presence - used to evaluate accuracy of predicted persistence, colonisation, extirpation

Focal Taxon

Focal Taxon: Vascular plants

Location

Location: Sweden

Scale of Analysis

Spatial extent: 11.57, 18.38, 56.16, 60.7 (xmin, xmax, ymin, ymax)

Spatial resolution: 4x4km

Temporal extent: 1929-2022

Temporal resolution: Two time periods: Historical period (1929-1970), modern period (2001-2022)

Boundary: political

Biodiversity data

Observation type: citizen science, field survey

Response data type: presence/absence

Predictors

Predictor types: climatic, edaphic, habitat

Hypotheses

Hypotheses: We hypothesize that species' occurrence predictions relate to their climatic niches (the average and range of temperatures experienced by each species over its range) and their habitat associations (proportions of each species' national populations restricted to grassland and forest habitat).

Assumptions

Model assumptions: Historical atlases had a lower recorder effort (often being the work of a single person), but we consider the presence data to be at least of a comparable quality to those used for species distribution modelling, with absences broadly likely to be true absences rather than non-detections.

Algorithms

Modelling techniques: GLM, GAM, FDA, GBM

Model complexity: We kept most of the model algorithm defaults. Our aim with this study was not to create the very best predictive models, but to build pretty good models according to a commonly-used method, in order to see how they match up to observed change.

Model averaging: Individual models –one per algorithm– of high predictive performance (true skill statistic TSS ≥0.7, or 10% of models with the highest TSS value) were ensembled following a committee averaging criteria to obtain a single model per species (Thuiller et al. 2009).

Workflow

Model workflow: In this study, we use historical (mid-1900s) climate, landscape and species occurrence data to build, and modern-day (1990-2020) data to evaluate species distribution models for 84 plant species in three regions of southern Sweden. After building our models based on the historical data, we use both traditional cross-validationory model evaluation measures within the same time period and comparison of changes in observed occurrence data over time to investigate the accuracy of the models in predicting change. To identify sources of uncertainty/error, Wwe then look into how species' occurrence predictions relate to their climatic niches (the average and range of temperatures experienced by each species over its range) and their habitat associations (proportions of each species' national populations restricted to grassland and forest habitat). Finally, we evaluate how prediction accuracy of extirpation, colonisation and persistence across all species is affected by the level amount of environmental change that has occurred within each grid-cell landscape over time.

Software

Software: BIOMOD2

Code availability: [redacted]

Data availability:]. For raw data, historical and modern species observations are available from the Swedish species observation system ArtPortalen (https://artportalen.se). Species temperature indices were taken from https://doi.org/10.6084/m9.figshare.8845832.v1 (Auffret and Thomas 2019), while habitat specialisation information was extracted from the supplementary information of Tyler et al. (2021). Climate data are free to access from https://www.smhi.se/data/ladda-ner-data/griddade-nederbord-och-temperaturdata-pthbv. Historical land-cover data are published at https://doi.org/10.17045/sthlmuni.4649854.v2 (Auffret et al. 2017), while modern land cover is published by the Swedish Environmental Protection Agency at https://www.naturvardsverket.se/verktyg-och-tjanster/kartor-och-karttjanster/nationella-marktackedata/ladda-ner-nationella-marktackedata/.

Data

Biodiversity data

Taxon names: 84 species of vascular plants: Allium ursinum, Alyssum alyssoides, Arctium lappa, Berberis vulgaris, Blysmus rufus, Botrychium lunaria, Brachypodium sylvaticum, Cakile maritima, Campanula latifolia, Cardamine amara, Cardamine bulbifera, Carex appropinquata, Carex distans, Carex pulicaris, Carex remota, Carex riparia, Catabrosa aquatica, Cerastium arvense, Chaenorhinum minus, Chenopodiastrum hybridum, Conium maculatum, Dactylorhiza incarnata, Daphne mezereum, Echium vulgare, Eleocharis parvula, Epilobium parviflorum, Epilobium roseum, Epilobium tetragonum, Epipactis helleborine, Equisetum hyemale, Equisetum pratense, Eriophorum latifolium, Filago arvensis, Galium mollugo, Geranium columbinum, Herminium monorchis, Honckenya peploides, Hottonia palustris, Isatis tinctoria, Jacobaea vulgaris, Jasione montana, Lathraea squamaria, Lathyrus niger, Lathyrus vernus, Lepidium ruderale, Leymus arenarius, Lithospermum officinale, Lonicera xylosteum, Matricaria chamomilla, Melica uniflora, Milium effusum, Monotropa hypopitys, Myrica gale, Myriophyllum alterniflorum, Myriophyllum spicatum, Neottia nidus-avis, Nepeta cataria, Ononis spinosa, Onopordum acanthium, Ophioglossum vulgatum, Plantago uniflora, Puccinellia distans, Ranunculus polyanthemos, Ribes alpinum, Rosa majalis, Rosa rubiginosa, Rubus sect. Corylifolii, Sanicula europaea, Schedonorus arundinaceus, Schedonorus giganteus, Sedum album, Sedum sexangulare, Sium latifolium, Sparganium erectum, Spergularia rubra, Stachys sylvatica, Tanacetum vulgare, Thelypteris palustris, Trifolium fragiferum, Utricularia vulgaris, Verbascum nigrum, Vincetoxicum hirundinaria, Viola mirabilis, Viola tricolor

Taxonomic reference system: the Swedish taxonomic database (https://namnochslaktskap.artfakta.se)

Ecological level: species

Data sources: Andersson, U.-B. and Gunnarsson, T. 2024. Ölands Flora. - SBF-förlag. Blomgren, E., Falk, E. and Herloff, B. 2011. Bohusläns Flora. - Föreningen Bohusläns Flora. Jonsell, L. 2010. Upplands Flora. - SBF Förlaget. Sterner, R. 1938. Flora der insel Öland. - Almqvist & Wiksells.

Sampling design: Plant occurrences used in this study originated from historical and modern-day regional plant biodiversity atlases (floras) from three Swedish provinces.

Sample size: Of the 84 vascular plant species, there were 26 409 recorded presences in 1055 grid squares in the historical period (counting each species only once per square), and 29 301 presences in 1074 grid squares in the historical period.

Clipping: Three Swedish provinces: Bohuslän, Uppland, Öland

Scaling: 4 × 4 km gridded climate data from the Swedish Institute for Hydrology and Meteorology's database pthbv, which we then aggregated to monthly values and resampled the grid to match the 5 × 5 km km Swedish national grid used for the species occurrence data.

Cleaning: NA

Absence data: We assume non-occurrences to equate to absences.

Background data: NA

Errors and biases: Unknown sampling effort

Data partitioning

Training data: We applied four runs per algorithm, to allow cross validation (each run selects a different random subset of 70% of data to fit the model).

Validation data: We applied four runs per algorithm, to allow cross validation (each run selects a different random subset of 70% of data to fit the model).

Test data: NA

Predictor variables

Predictor variables: Climate variables: Temperature (°C): Mean annual temperature; Mean of monthly maximum temperatures; Mean of monthly minimum temperatures. Precipitation (mm): Total annual precipitation; Mean of monthly minimum precipitation, Mean of monthly maximum precipitation. Source: Swedish Meteorological and Hydrological Institute database of gridded daily values (4 km resolution, resampled to 5 × 5 km grid). Historical period 1961-1970; Modern period 2001-2010.

Land use variables: Fraction cover of grassland, forest, arable fields and surface water. Source: Historical data from Economic maps published 1940s-1960s (1 m resolution, aggregated and resampled to 10 m; Auffret et al. 2017a). Grassland category mainly consists of grasslands, but also includes other open land cover, including wetlands. Because of difficulties in separating water and forest in the digitisation process, lakes and rivers were added to the digitisations using a modern map layer. Modern data from 2018 Swedish land-cover map (10 m resolution), with 25 categories re-classified to match the broad categories of the historical data.

Soil variables: Fraction of peat, sand & gravel, bedrock, silt & clay, till and other soils. Source: National soil maps. Original data in vector format at 1:25 000 to 1:1 000 000 scales were harmonised to create a 50 m resolution layer (noting that the smallest features on the coarsest map are approximately 500 m across).

Distance to coast: Euclidean distance between the grid cell centroid and the Swedish coast.

Data sources: Climate data are free to access from https://www.smhi.se/data/ladda-ner-data/griddade-nederbord-och-temperaturdata-pthbv. Historical land-cover data are published at https://doi.org/10.17045/sthlmuni.4649854.v2 (Auffret et al. 2017), while modern land cover is published by the Swedish Environmental Protection Agency at https://www.naturvardsverket.se/verktyg-och-tjanster/kartor-och-karttjanster/nationella-marktackedata/ladda-ner-nationella-marktackedata/.

Spatial extent: 11.57, 18.38, 56.16, 60.7 (xmin, xmax, ymin, ymax)

Spatial resolution: The atlas data is point data

Coordinate reference system: EPSG: 3006

Temporal extent: 1929-2022

Temporal resolution: Each time period contains data covering 20-approx 40 years.

Data processing: Land-use aggregated to four categories: grassland, forest, arable fields and surface water

Errors and biases: NA

Dimension reduction: See next section.

Transfer data

Data sources: For raw data, historical and modern species observations are available from the Swedish species observation system ArtPortalen (https://artportalen.se). Climate data are free to access from https://www.smhi.se/data/ladda-ner-data/griddade-nederbord-och-temperaturdata-pthbv. Historical land-cover data are published at https://doi.org/10.17045/sthlmuni.4649854.v2 (Auffret et al. 2017), while modern land cover is published by the Swedish Environmental Protection Agency at https://www.naturvardsverket.se/verktyg-och-tjanster/kartor-och-karttjanster/nationella-marktackedata/ladda-ner-nationella-marktackedata/.

Spatial extent: 11.57, 18.38, 56.16, 60.7 (xmin, xmax, ymin, ymax)

Spatial resolution: Climate data: 4 × 4 km gridded climate data from the Swedish Institute for Hydrology and Meteorology's database pthbv, which we then aggregated to monthly values and resampled the grid to match the 5 × 5 km km Swedish national grid used for the species occurrence data. Land-use data: 10m, summarised to 5km.

Temporal extent: Climate data: 2001-2010; Land use data: 2018.

Temporal resolution: NA

Models and scenarios: We project into the Modern time period.

Data processing: NA

Quantification of Novelty: We conducted a multivariate similarity surface (MESS) analysis to identify where novel environments exist in the modern data, and which predictor variables are responsible. We used the MESS implementation of the ecospat package (Di Cola et al. 2017), on the variables that changed over time and that were included in the SDMs: maximum annual temperature and precipitation, fractions of arable, open and forest land.

Model

Variable pre-selection

Variable pre-selection: Variables that were not included in final models due to multicollinearity were: Mean annual temperature, Mean of monthly minimum temperatures, Total annual precipitation, Mean of monthly minimum precipitation, Cover of surface water, Proportion of till soils.

Multicollinearity

Multicollinearity: Only non-correlated environmental predictors (VIF≤4) were kept for the modelling, including maximum annual temperature and precipitation, fractions of arable, open and forest land (all from the historical period), as well as distance to the coast, and proportions of peat, sand/gravel, rock and silt/clay soil types.

Model settings

GLM: response (quadratic), Interaction.level (0)

GAM: smoothing term (k), (3)

FDA: all (default)

GBM: all (default)

Model settings (extrapolation): We only predicted within the three counties from which the state data were collected.

Model estimates

Coefficients: NA

Parameter uncertainty: NA

Variable importance: NA

Model selection - model averaging - ensembles

Model selection: NA

Model averaging: NA

Model ensembles: NA

Analysis and Correction of non-independence

Spatial autocorrelation: Spatial autocorrelation was controlled for by including as predictors the first two eigenvectors of an analysis of the principal coordinates of neighbour matrices, based on the centroids of each grid cell (Borcard and Legendre 2002).

Temporal autocorrelation: NA

Nested data: NA

Threshold selection

Threshold selection: Predictions were converted into a binary classification using a threshold value (specific for that species and model) that was automatically selected to maximise the model's TSS score, based on the 30% subset of observations not used to build each candidate model (Liu et al. 2013). Choice of threshold value for converting probability of occurrence to a binary measure is a source of variation in predictions of turnover at the grid-cell level (Nenzén and Araújo 2011). Therefore, we performed a sensitivity analysis in which we re-ran the above analyses of prediction accuracy at both the species level and grid-cell level, but with a lower (TSS-chosen threshold minus 0.1 on the 0-1 scale) and higher (TSS-chosen threshold plus 0.1) threshold for converting probability of occurrence into predicted presence from the ensemble forecasts.

Assessment

Performance statistics

Performance on training data: AUC, TSS

Performance on validation data: AUC, TSS

Performance on test data: AUC, TSS

Plausibility check

Response shapes: NA

Expert judgement: NA

Prediction

Prediction output

Prediction unit: 5 x 5 km grid cells

Post-processing: NA

Uncertainty quantification

Algorithmic uncertainty: NA

Input data uncertainty: NA

Parameter uncertainty: NA

Scenario uncertainty: We predict to the Modern period (2001-2022) for which we have observed data (both environmental and biodiversity).

Novel environments: We conducted a multivariate similarity surface (MESS) analysis to identify where novel environments exist in the modern data, and which predictor variables are responsible. We used the MESS implementation of the ecospat package (Di Cola et al. 2017), on the variables that changed over time and that were included in the SDMs: maximum annual temperature and precipitation, fractions of arable, open and forest land.