Ecography

ECOG-03888
Chakraborty, D., Schueler, S., Lexer, M. and Wang, T. 2018. Genetic trials improve the transfer of Douglas-fir distribution models across continents. – Ecography doi: 10.1111/ecog.03888

Supplementary material

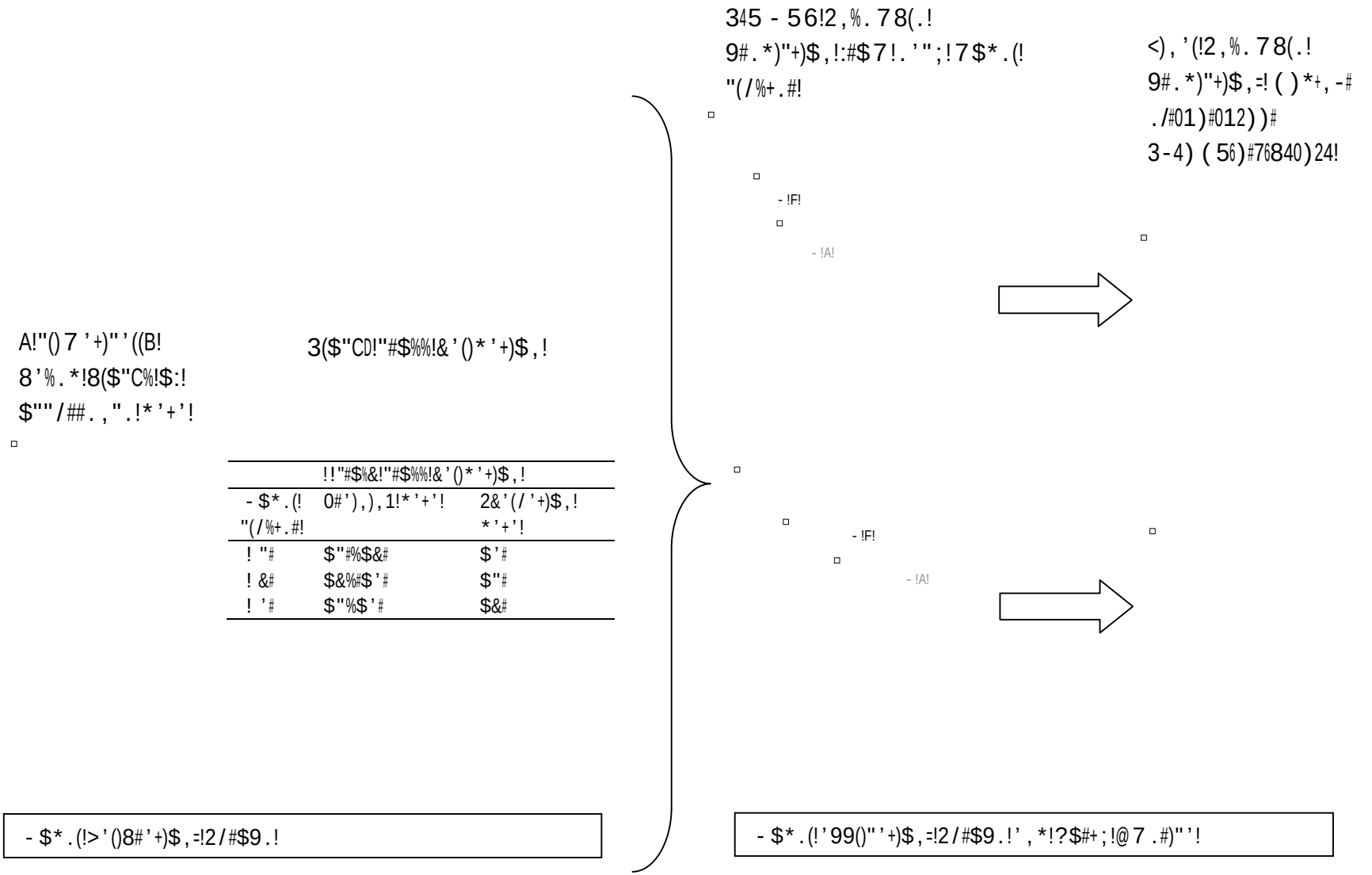

Fig. A3 Schematic work flow of the *occurrence-based SDM* model calibration within the BIOMOD 2 platform. The occurrence data (presence+ absence of Douglas-fir in Europe) was clustered into 3 climatically similar blocks. *K-fold* cross validation was done with this clustered dataset using two blocks for model training and with leaving one out for model validation . Each of the three model cluster predicted probability of occurrence of Douglas-fir in Europe and North America with the10 models available in the BIOMOD 2 platform. Ensemble predictions were then obtained from each of the three model clusters for both Europe and North America. A final ensemble model (referred to in the paper as "correlative SDM") prediction was obtained with a median of the three ensemble clusters, each for Europe and North America.

Table A1. List of climate variables screened to developed the Universal Response Function (Chakraborty *et al.*, 2015) and the occurrence-based SDM.

Climate variable	Acronym
Mean annual temperature (°C),	MAT
Mean warmest month temperature (°C),	MWMT
Mean coldest month temperature (°C),	MCMT
Temperature difference between MWMT and MCMT, or continentality (°C),	TD
Mean annual precipitation (mm),	MAP
Mean summer (May to Sept.) precipitation (mm),	MSP
Annual heat: moisture index (MAT+10)/(MAP/1000))	AHM
Summer heat: moisture index ((MWMT)/(MSP/1000))	SHM
Degree-days below 0°C, chilling degree-days	DD<0
Degree-days above 5°C, growing degree-days	DD>5
Degree-days below 18°C, heating degree-days	DD<18
Degree-days above 18°C, cooling degree-days	DD>18
The number of frost-free days	NFFD
Frost-free period	FFP
The Julian date on which FFP begins	bFFP
The Julian date on which FFP ends	eFFP
Precipitation as snow (mm) between August in previous year and July in current year	PAS
Extreme minimum temperature over 30 years	EMT
Hargreaves reference evaporation	Eref
Hargreaves climatic moisture deficit	CMD

Table A2. Relative importance of the climate variables used to predict the distribution of Douglas-fir in Europe obtained from ten different modeling algorithms and the ensemble model of the BIOMOD2 platform. Relative importance range from 0-1shown with red to green colors with 0 (red cells) having very little or no influence and 1(green cells) having highest influence on the model. For explanation of the climate variables see Table A1

Model cluster	Variable	GLM	GAM	GBM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT	Ensemble
M1	DDabove5	0.44	0.38	0.43	0.49	0.34	0.21	0.43	0.41	0.75	0.41	0.43
M1	Eref	0.11	0.05	0.29	0.41	0.06	0.26	0.29	0.20	0.12	0.04	0.18
M1	SHM	0.09	0.25	0.01	0.17	0.05	0.14	0.07	0.08	0.27	0.07	.12
M1	MAT	1.00	1.00	1.00	0.47	1.00	0.25	1.00	1.00	0.54	1.00	0.83
M1	TD	0.29	0.84	0.12	0.21	0.53	0.27	0.00	0.20	0.17	0.42	0.30
M1	PAS	0.17	0.06	0.08	0.00	0.04	0.22	0.08	0.25	0.13	0.04	0.11
M1	MCMT	0.31	0.51	0.11	0.11	0.32	0.26	0.18	0.31	0.16	0.35	0.26
M2	DDabove5	0.79	0.73	0.78	0.65	0.42	0.51	0.89	0.00	0.75	0.74	0.62
M2	Eref	0.00	0.00	0.89	0.77	0.82	0.60	1.00	1.00	0.12	0.10	0.53
M2	SHM	0.62	0.81	0.65	0.68	0.87	0.22	0.73	0.57	0.27	0.74	0.62
M2	MAT	1.00	0.99	0.46	0.46	0.49	0.51	0.68	0.55	0.54	1.00	0.67
M2	TD	0.00	0.00	0.87	0.59	0.07	0.25	0.87	0.98	0.17	0.28	0.41
M2	PAS	0.72	0.69	0.12	0.22	0.41	0.23	0.02	0.31	0.13	0.60	0.34
M2	MCMT	0.66	0.00	0.56	0.16	0.63	0.37	0.56	0.15	0.16	0.53	0.38
M3	DDabove5	0.73	0.44	0.89	0.56	0.76	0.16	0.89	0.69	0.69	0.82	0.66
M3	Eref	0.08	0.11	0.16	0.48	0.15	0.22	0.11	0.04	0.12	0.04	0.15
M3	SHM	0.39	0.02	0.49	0.38	0.58	0.22	0.49	0.39	0.37	0.40	0.37
M3	MAT	0.85	1.00	1.00	0.44	0.70	0.21	1.00	0.93	0.54	0.89	0.76
M3	TD	0.85	0.30	0.56	0.16	0.29	0.20	0.49	0.55	0.21	0.60	0.42
M3	PAS	0.06	0.17	0.02	0.00	0.11	0.16	0.00	0.00	0.16	0.01	0.07
M3	MCMT	0.78	0.31	0.80	0.23	0.35	0.21	0.80	0.58	0.16	0.66	0.49

Table A3. Results of block cross- validation of the ten modeling algorithm and the ensemble model for evaluating the model clusters with occurrence data of Douglas-fir in Europe. For details see Fig. A2

Model clusters	Models	Training data	Evaluation data	TSS in Testing data	TSS in Evaluation data	Sensitivity	Specificity
M1	GLM	C1+C2	C3	0.99	0.59	0.86	0.73
M1	GAM	C1+C2	C3	0.99	0.45	0.86	0.59
M1	GBM	C1+C2	C3	0.99	0.53	0.64	0.89
M1	CTA	C1+C2	C3	0.99	0.45	0.86	0.59
M1	ANN	C1+C2	C3	0.99	0.88	0.99	0.89
M1	SRE	C1+C2	C3	0.89	-0.01	0.99	0.00
M1	FDA	C1+C2	C3	0.99	0.96	0.99	0.97
M1	MARS	C1+C2	C3	0.98	0.90	0.99	0.91
M1	RF	C1+C2	C3	0.98	0.86	0.99	0.87
M1	MAXENT	C1+C2	C3	0.98	0.85	0.98	0.87
M1	Ensemble mean	C1+C2	C3	0.98	0.71	0.98	0.73
M1	Ensemble median	C1+C2	C3	0.99	0.85	0.98	0.87
M2	GLM	C2+C3	C1	0.97	0.97	0.98	0.99
M2	GAM	C2+C3	C1	0.94	0.95	0.98	0.97
M2	GBM	C2+C3	C1	0.99	0.73	0.98	0.75
M2	CTA	C2+C3	C1	0.95	0.91	0.98	0.93
M2	ANN	C2+C3	C1	0.90	0.93	0.98	0.95
M2	SRE	C2+C3	C1	0.82	-0.02	0.98	0.00
M2	FDA	C2+C3	C1	0.90	0.90	0.98	0.92
M2	MARS	C2+C3	C1	0.96	0.89	0.98	0.91
M2	RF	C2+C3	C1	1.00	0.75	0.98	0.77
M2	MAXENT	C2+C3	C1	0.94	0.96	0.98	0.98
M2	Ensemble mean	C2+C3	C1	0.94	0.80	0.98	0.82
M2	Ensemble median	C2+C3	C1	0.94	0.90	0.98	0.92
M3	GLM	C1+C3	C2	0.99	0.97	0.98	0.99
M3	GAM	C1+C3	C2	0.99	0.87	0.98	0.89
M3	GBM	C1+C3	C2	1.00	0.97	0.98	0.99
M3	CTA	C1+C3	C2	0.99	0.63	0.98	0.65
M3	ANN	C1+C3	C2	0.99	0.98	0.98	1.00
M3	SRE	C1+C3	C2	0.88	-0.02	0.98	0.00
M3	FDA	C1+C3	C2	0.99	0.44	0.44	1.00
M3	MARS	C1+C3	C2	0.98	0.55	0.56	1.00
M3	RF	C1+C3	C2	0.98	0.55	0.56	0.99
M3	MAXENT	C1+C3	C2	0.98	0.54	0.56	0.99
M3	Ensemble mean	C1+C3	C2	0.98	0.38	0.63	0.75
M3	Ensemble median	C1+C3	C2	0.99	0.55	0.56	0.99

Sensitivity = the proportion of true presences correctly identified by the model
Specificity = the proportion of true absences correctly identified by the model
True Skill Statistics = measure of model performance independent of prevalence. TSS value range from -1 to +1 where +1 indicate perfect agreement and values <=0 indicate a performance equivalent to a random guess.

Table. A4 Evaluation of the ten modeling algorithm and the ensemble model within each model cluster, for predicting the distribution of Douglas-fir in Europe and North America.

Model cluster	Models	Application in Europe			Application in North America		
		TSS	Sensitivity	Specificity	TSS	Sensitivity	Specificity
M1	GLM	0.96	0.99	0.97	-0.076	0.11	0.82
M1	GAM	0.98	0.98	0.99	0.046	0.09	0.96
M1	GBM	0.98	0.98	0.99	0.077	0.16	0.92
M1	CTA	0.91	0.99	0.92	0.163	0.38	0.78
M1	ANN	0.98	0.99	0.99	0.046	0.09	0.96
M1	SRE	0.78	0.80	0.97	0.025	0.03	0.99
M1	FDA	0.96	0.99	0.97	0.007	0.10	0.90
M1	MARS	0.98	0.98	0.99	-0.048	0.08	0.87
M1	RF	0.99	0.99	0.99	0.085	0.17	0.92
M1	MAXENT	0.97	0.99	0.99	0.055	0.10	0.95
M1	Ensemble mean	0.98	0.99	0.99	0.038	0.13	0.91
M1	Ensemble median	0.98	0.99	0.99	0.046	0.10	0.92
M2	GLM	0.09	0.35	0.74	0.329	0.51	0.82
M2	GAM	0.86	0.93	0.93	0.045	0.12	0.92
M2	GBM	0.69	0.71	0.98	0.149	0.24	0.91
M2	CTA	0.63	0.99	0.64	0.210	0.91	0.30
M2	ANN	0.86	0.93	0.93	0.174	0.26	0.91
M2	SRE	0.41	0.76	0.65	0.239	0.32	0.92
M2	FDA	0.90	0.95	0.95	0.210	0.60	0.61
M2	MARS	0.80	0.99	0.81	0.253	0.88	0.37
M2	RF	0.70	0.72	0.98	0.169	0.28	0.89
M2	MAXENT	0.68	0.99	0.68	0.050	0.12	0.93
M2	Ensemble mean	0.78	0.99	0.79	0.183	0.43	0.76
M2	Ensemble median	0.65	0.99	0.66	0.192	0.30	0.90
M3	GLM	0.98	0.99	0.99	-0.064	0.13	0.81
M3	GAM	0.97	0.99	0.98	0.210	0.60	0.00
M3	GBM	0.98	0.99	0.99	0.067	0.15	0.91
M3	CTA	0.97	0.99	0.99	0.053	0.17	0.88
M3	ANN	0.98	0.99	0.99	0.036	0.10	0.94
M3	SRE	0.78	0.80	0.98	0.025	0.03	1.00
M3	FDA	0.97	0.99	0.98	0.041	0.12	0.92
M3	MARS	0.98	0.99	0.99	0.035	0.11	0.93
M3	RF	0.98	0.99	0.99	0.070	0.16	0.91
M3	MAXENT	0.96	0.97	0.99	0.054	0.10	0.95
M3	Ensemble mean	0.98	0.99	0.99	0.035	0.12	0.92
M3	Ensemble median	0.98	0.99	0.99	0.041	0.12	0.92

Sensitivity = the proportion of true presences correctly identified by the model
Specificity = the proportion of true absences correctly identified by the model
True Skill Statistics = measure of model performance independent of prevalence. TSS value range from -1 to +1 where +1 indicate perfect agreement and values <=0 indicate a performance equivalent to a random guess.

Table A5. The Universal response Functions (URFs) with only the Coastal provenances. Results of multiple regression analysis predicting dominant height at age 24 and basal area at age 24 of Douglas-fir populations from the site and seed origin climate as independent variables in an URF.

Independent variables	**Parameter estimate**	**St error**	**95% Confidence interval**		***p-value***
URF for dominant tree height (H24) [m]					
Intercept	-18.730	0.298	-19.217	-18.236	<0.001
MAT_s	6.233	0.068	6.121	6.346	<0.001
MAT_s^2	-0.467	0.005	-0.475	-0.459	<0.001
SHM_s	0.480	0.012	0.460	0.500	<0.001
SHM_s^2	-0.005	0.000	-0.005	-0.005	<0.001
MAT_p	1.186	0.038	1.124	1.249	<0.001
MAT_p^2	-0.122	0.002	-0.126	-0.119	<0.001
MAT_s*MAT_p	0.092	0.004	0.086	0.098	<0.001
Model R^2adj				**0.92**	
URF for basal area (BA24) [m^2ha^{-1}]					
Intercept	-41.71	1.02	-43.39	-40.03	<0.001
MAT_s	10.79	0.24	10.41	11.18	<0.001
MAT_s^2	-0.63	0.02	-0.66	-0.60	<0.001
SHM_s	0.56	0.04	0.49	0.63	<0.001
SHM_s^2	-0.01	0.00	-0.01	-0.01	<0.001
MAT_p	3.84	0.13	3.63	4.05	<0.001
MAT_p^2	-0.24	0.01	-0.26	-0.23	<0.001
MAT_s*MAT_p	-0.02	0.01	-0.04	0.00	<0.001
Model R^2adj				**0.88**	

MAT_s = Mean annual temperature of planting site
SHM_s = Summer heat moisture Index (Mean temperature of Warmest month/ (Mean summer temperature (May-Sep)/1000))
MAT_p = mean annual temperature of seed origin

Table A6. Performance and transferability of the occurrence-based SDM and the two trait based URF SDMs calibrated with only coastal provenances of Douglas-fir. The performance of the three SDMs for predicting observed presence/absence of Douglas-fir in its and introduced range in Europe (EU) and its native range in North America (NA), and in the coastal distribution of Douglas-fir (Coastal NA).

Type of SDM	**Developed in**	**Applied in**	**Sensitivity**	**Specificity**	**TSS**	**T_{Tss}**
URF-SDM for H24	EU	EU	0.98	0.54	0.52	
		NA	0.65	0.71	0.36	0.69
		Coastal NA	0.62	0.58	0.20	0.38
URF-SDM for BA24	EU	EU	0.99	0.53	0.53	
		NA	0.70	0.69	0.38	0.73
		Coastal NA	0.60	0.59	0.19	0.36
Occurrence-based SDM	EU	EU	0.99*	0.99*	0.98*	
		NA	0.12*	0.92*	0.05*	0.05
		Coastal NA	0.43*	0.59*	0.02*	0.02

Sensitivity = the proportion of true presences correctly identified by the model
Specificity = the proportion of true absences correctly identified by the model
True Skill Statistics = measure of model performance independent of prevalence. TSS value range from -1 to +1 where +1 indicate perfect agreement and values <=0 indicate a performance equivalent to a random guess.
T_{Tss} = Transferability Index (Heikkinen *et al.*, 2012). T_{TSS} <1 provides higher predictive accuracy in interpolative forecasting than in extrapolation, and vice versa
* median sensitivity, specificity and TSS of the three ensemble model clusters (See Fig. A2 for details).