Appendix A

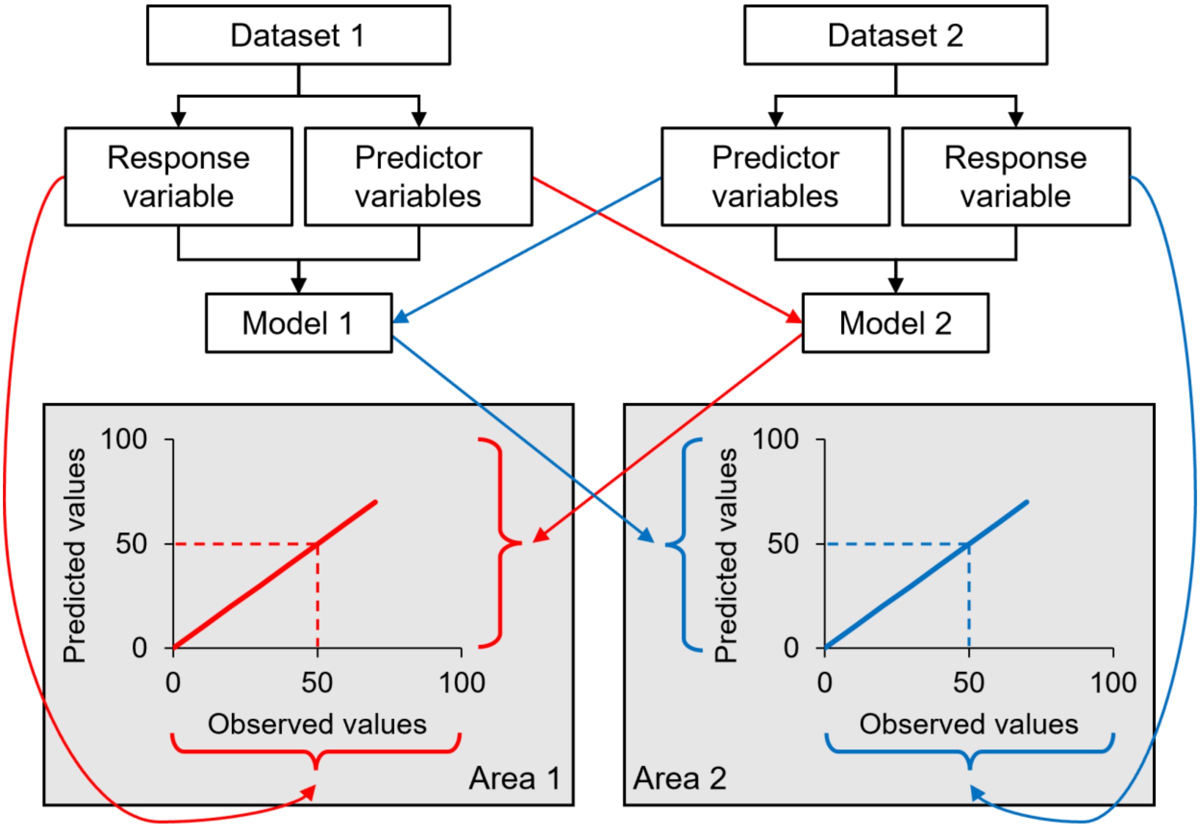

Fig. A1. Testing model transferability. The transferability of Model 1 (calibrated in Area 1; blue lines) would be tested by predicting the response variable in Area 2 using the predictor variables from Dataset 2. These predicted values would be compared to the observed values in Dataset 2. The opposite can be done (red lines) to test the transferability of Model 2. The graphs in each area display the ideal one-to-one relationship between predicted and observed values. Figure adapted from Randin et al. (2006).

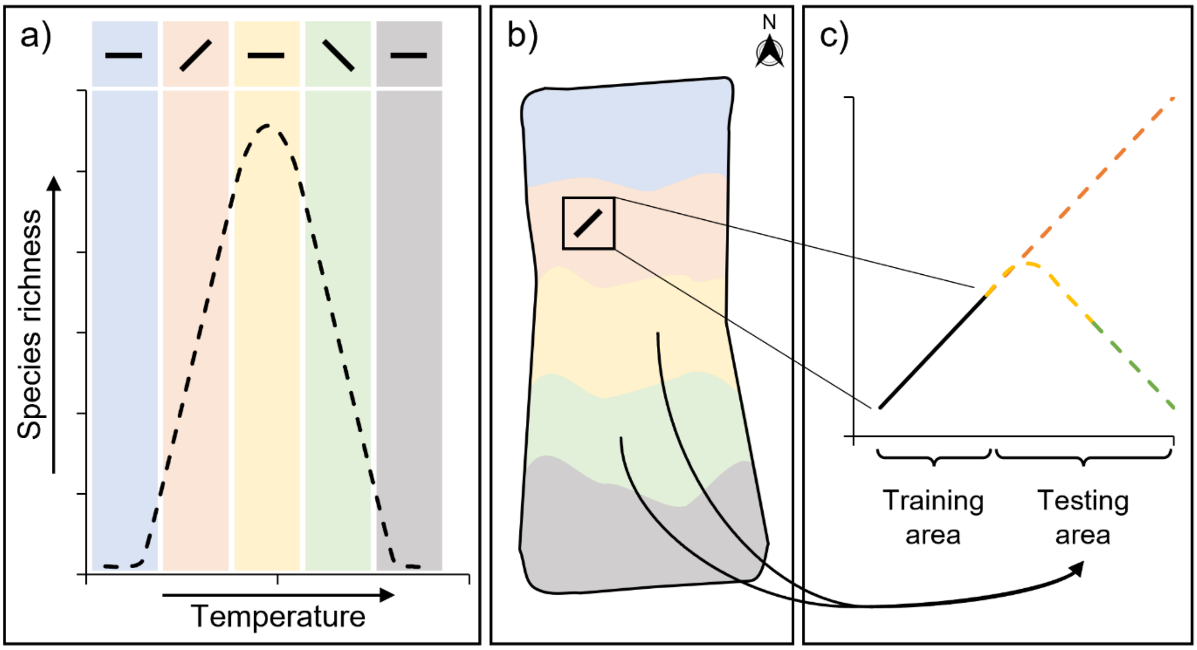

Fig. A2. A diagram displaying the issues of novel environmental conditions. a) A hypothetical plot of the response of species richness to an environmental factor, such as temperature. In this diagram shading correspond to different temperature values. b) A map of a hypothetical area showing a gradient of lower temperatures in the north and higher temperatures in the south. If, for example, only the area within the training area (represented by the quadrat) is sampled, then the relationship between species richness and temperature will appear to be a positive linear relationship. c) If the results from the training area are used to extrapolate species richness to other areas the model will predict the positive linear relationship (orange dashed line), instead of the more accurate relationship (yellow and green dashed line). Figure adapted from Rousseau and Betts (2022).

References

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33, 1689–1703.

Rousseau, J.S., Betts, M.G., 2022. Factors influencing transferability in species distribution models. Ecography e06060, 1–13.

Appendix B

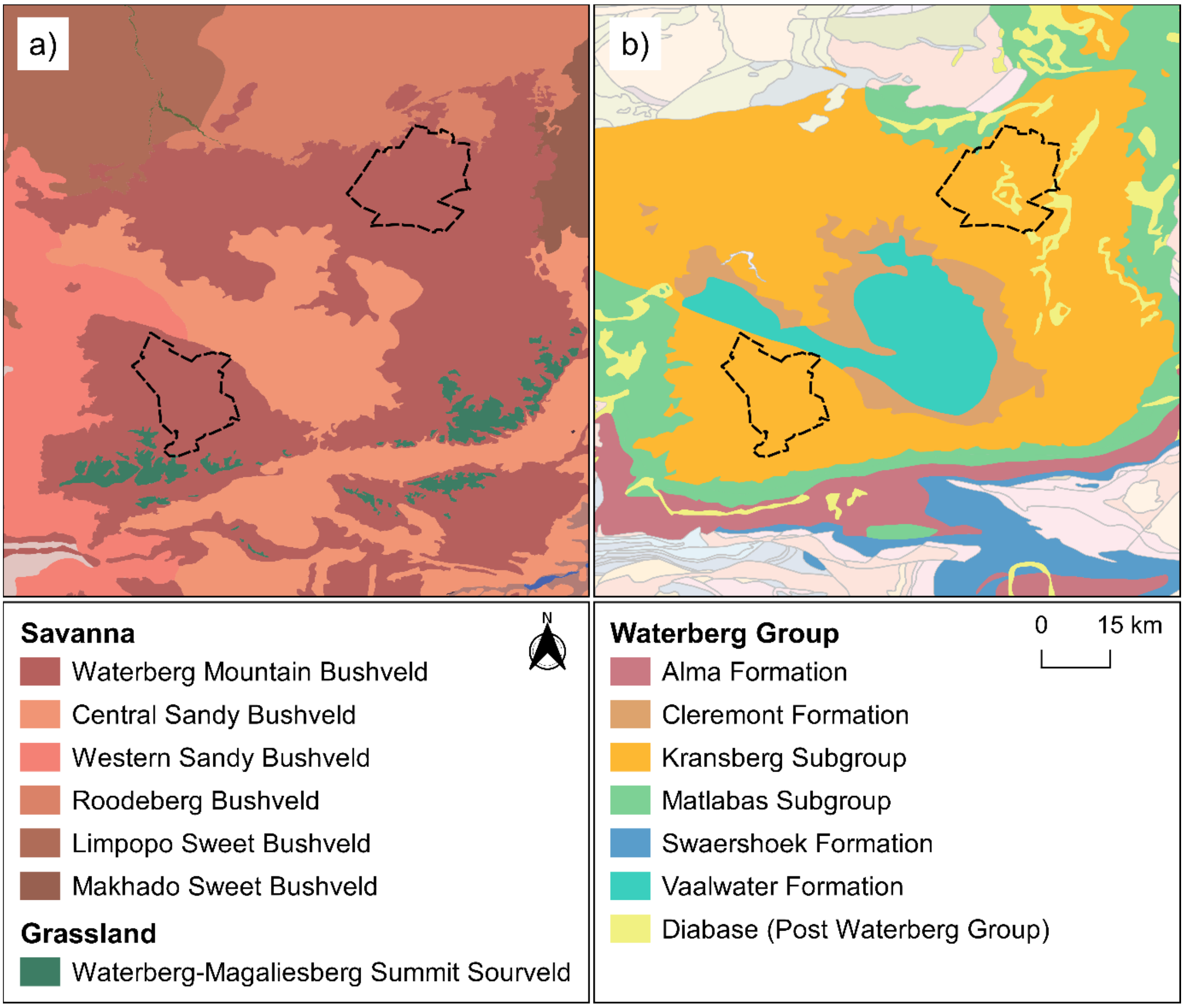

Fig. B1. The a) vegetation (Mucina and Rutherford, 2006) and b) underlying geology (Council for Geoscience, 2020) of the two game reserves (dashed black lines), Lapalala (north-east) and Welgevonden (south-west).

References

Council for Geoscience, 2020. 1:1 000 000 Geological Data [WWW Document]. URL https://www.geoscience.org.za/index.php/publication/downloadable-material (accessed 5.25.20).

Mucina, L., Rutherford, M.C., 2006. The Vegetation of South Africa, Lesotho and Swaziland. *Strelitzia* 19. South African National Biodiversity Institute, Pretoria.

Appendix C

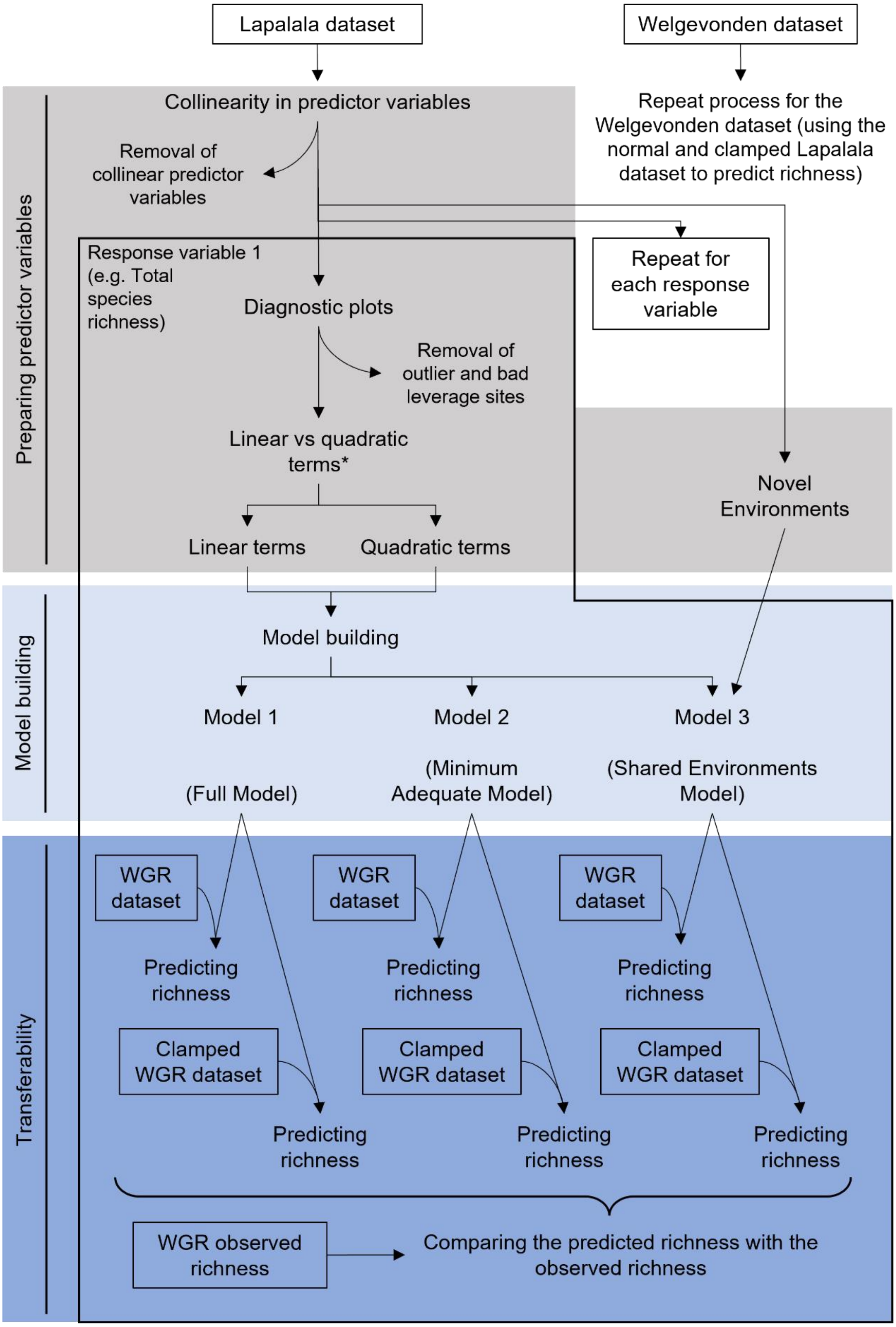

Fig. C1. The steps followed in the data analyses section for this study's methods. WGR = Welgevonden. * The linear vs quadratic terms step was part of the generalised boosted model building.

Appendix D

Table D1

A summary comparing the results of the linear (LM) and quadratic (QM) univariate models for each predictor variable against each of the six response variables for both the Lapalala and Welgevonden datasets. The change in AIC value and the change in deviance explained (DE) was used to determine which of the two models performed better. A negative change in AIC value (meaning the quadratic model's AIC value is lower than the linear model's AIC value) and a 5% or greater change in DE lead to the quadratic model being classified as the better model.

		Lapalala				Welgevonden			
Response	Predictor	Change in AIC	LM DE (%)	Change in DE (%)	Better model	Change in AIC	LM DE (%)	Change in DE (%)	Better model
Total species richness	Elevation	2	0.49	0.01	LM	-5	1.33	5.03	QM
	Slope	1	0.18	0.21	LM	1	2.26	0.78	LM
	Curvature	1	0.01	0.10	LM	0	3.77	1.65	LM
	TWI	1	1.01	0.10	LM	1	6.19	0.51	LM
	PDIR	-1	0.04	0.44	LM	1	0.33	0.98	LM
	Rock cover	-27	5.41	4.98	LM	2	2.89	0.04	LM
	Woody debris cover	0	0.25	0.41	LM	-14	10.41	11.15	QM
	Bare soil cover	-1	4.65	0.46	LM	-2	0.01	2.79	LM
	Leaf litter cover	-9	6.68	1.91	LM	-2	2.78	2.73	LM
	Soil pH	1	8.04	0.13	LM	-13	0.00	10.78	QM
	Clay	0	0.18	0.34	LM	1	3.14	0.65	LM
	Carbon	-14	14.08	2.80	LM	-8	1.15	6.72	QM
	Potassium	-26	1.29	4.82	LM	0	2.94	1.40	LM
	Calcium	-10	6.47	2.12	LM	-2	4.26	2.96	LM
	Sodium	1	2.89	0.18	LM	1	6.21	0.79	LM
	Phosphorus	-30	4.21	5.41	QM	1	1.39	0.43	LM
Grass species richness	Elevation	2	0.03	0.09	LM	1	0.48	2.19	LM
	Slope	0	0.13	1.13	LM	2	2.87	0.76	LM
	Curvature	2	0.03	0.27	LM	2	0.01	0.69	LM
	TWI	1	1.43	0.36	LM	1	4.30	1.50	LM
	PDIR	1	0.00	0.53	LM	1	0.12	2.30	LM
	Rock cover	0	0.03	0.93	LM	2	11.37	0.46	LM
	Woody debris cover	-1	1.06	1.82	LM	2	0.57	0.01	LM
	Bare soil cover	1	2.08	0.76	LM	2	0.28	0.74	LM
	Leaf litter cover	-3	0.00	2.78	LM	1	0.25	1.41	LM
	Soil pH	2	9.46	0.08	LM	-5	2.07	12.65	QM
	Clay	-4	1.86	3.68	LM	0	4.92	3.54	LM
	Carbon	0	3.37	1.31	LM	2	13.49	0.02	LM
	Potassium	2	9.42	0.02	LM	2	5.64	0.05	LM
	Calcium	1	10.07	0.67	LM	-3	0.73	7.81	QM
	Sodium	1	5.92	0.45	LM	1	2.11	1.58	LM
	Phosphorus	2	0.25	0.04	LM	1	13.20	2.00	LM

Table D1. Continued.

Response	Predictor	Lapalala				Welgevonden			
		Change in AIC	LM DE (%)	Change in DE (%)	Better model	Change in AIC	LM DE (%)	Change in DE (%)	Better model
Herb species richness	Elevation	-1	0.85	0.53	LM	-2	3.71	3.09	LM
	Slope	2	0.93	0.06	LM	1	0.40	1.00	LM
	Curvature	2	0.00	0.00	LM	1	2.52	0.88	LM
	TWI	-1	0.02	0.60	LM	-2	7.13	3.15	LM
	PDIR	2	1.79	0.08	LM	2	0.45	0.06	LM
	Rock cover	-15	1.13	3.12	LM	2	6.48	0.15	LM
	Woody debris cover	-8	0.49	1.90	LM	-4	25.60	4.63	LM
	Bare soil cover	-8	0.03	1.73	LM	0	0.66	1.92	LM
	Leaf litter cover	-21	0.15	4.25	LM	-3	1.57	3.76	LM
	Soil pH	0	0.16	0.31	LM	-13	0.28	12.75	QM
	Clay	1	0.00	0.27	LM	2	2.12	0.21	LM
	Carbon	-3	0.08	0.94	LM	-5	0.65	5.98	QM
	Potassium	-11	1.87	2.36	LM	2	3.48	0.11	LM
	Calcium	1	0.02	0.18	LM	2	2.41	0.41	LM
	Sodium	-2	0.36	0.72	LM	2	6.40	0.06	LM
	Phosphorus	-31	1.70	6.05	QM	-1	0.71	2.25	LM
Woody species richness	Elevation	-1	2.21	0.84	LM	2	7.42	0.05	LM
	Slope	2	4.68	0.04	LM	1	9.25	0.31	LM
	Curvature	1	1.01	0.16	LM	0	0.18	0.96	LM
	TWI	0	1.11	0.60	LM	0	0.16	0.83	LM
	PDIR	-9	0.02	2.73	LM	-17	0.02	8.80	QM
	Rock cover	-16	12.74	4.32	LM	-4	21.50	2.99	LM
	Woody debris cover	-4	2.41	1.41	LM	-36	30.61	17.33	QM
	Bare soil cover	2	8.98	0.00	LM	-5	0.83	3.05	LM
	Leaf litter cover	-6	23.28	1.85	LM	-4	23.05	2.56	LM
	Soil pH	1	11.77	0.20	LM	2	6.25	0.15	LM
	Clay	-1	0.06	0.74	LM	0	3.92	1.08	LM
	Carbon	-23	31.58	6.02	QM	-10	25.16	5.33	QM
	Potassium	-13	4.74	3.62	LM	1	0.96	0.61	LM
	Calcium	-28	17.75	7.18	QM	-2	12.65	1.69	LM
	Sodium	-3	3.79	1.33	LM	-14	3.17	7.41	QM
	Phosphorus	-23	1.00	6.06	QM	-10	4.16	5.49	QM

Table D1. Continued.

Response	Predictor	Lapalala				Welgevonden			
		Change in AIC	LM DE (%)	Change in DE (%)	Better model	Change in AIC	LM DE (%)	Change in DE (%)	Better model
Family richness	Elevation	1	0.14	0.23	LM	1	2.35	1.51	LM
	Slope	2	3.92	0.00	LM	2	7.65	0.43	LM
	Curvature	-1	0.01	1.00	LM	2	1.54	0.05	LM
	TWI	2	0.16	0.00	LM	1	3.35	2.01	LM
	PDIR	1	0.20	0.51	LM	-2	0.28	6.00	QM
	Rock over	-5	12.85	2.83	LM	0	15.69	3.16	LM
	Woody debris cover	0	0.02	0.65	LM	-4	15.14	9.06	QM
	Bare soil cover	2	7.46	0.08	LM	1	1.01	1.86	LM
	Leaf litter cover	-2	6.88	1.75	LM	0	2.59	3.22	LM
	Soil pH	2	8.21	0.16	LM	1	0.80	2.14	LM
	Clay	1	0.06	0.29	LM	0	0.19	2.76	LM
	Carbon	-5	26.92	2.79	LM	-8	14.83	13.88	QM
	Potassium	-13	2.52	5.81	QM	2	2.96	0.17	LM
	Calcium	-4	11.35	2.39	LM	2	9.55	0.02	LM
	Sodium	0	4.30	0.67	LM	0	9.42	2.96	LM
	Phosphorus	-14	5.85	6.33	QM	2	12.83	0.69	LM
Genus richness	Elevation	-2	0.05	0.68	LM	-4	5.09	3.96	LM
	Slope	1	0.44	0.21	LM	1	4.43	0.37	LM
	Curvature	2	0.02	0.09	LM	2	3.08	0.00	LM
	TWI	1	0.43	0.16	LM	0	4.05	1.18	LM
	PDIR	0	0.01	0.42	LM	-1	1.27	2.31	LM
	Rock cover	-17	5.54	3.66	LM	2	4.68	0.04	LM
	Woody debris cover	-4	0.10	1.13	LM	-18	13.82	13.62	QM
	Bare soil cover	2	6.85	0.03	LM	-2	0.15	2.93	LM
	Leaf litter cover	-9	3.27	2.10	LM	-1	4.62	2.33	LM
	Soil pH	1	10.76	0.23	LM	-10	0.02	7.95	QM
	Clay	1	0.54	0.29	LM	1	4.39	0.36	LM
	Carbon	-15	16.53	3.30	LM	-9	1.85	7.52	QM
	Potassium	-11	4.57	2.52	LM	1	2.18	0.80	LM
	Calcium	-16	8.95	3.58	LM	-1	4.52	2.00	LM
	Sodium	2	5.44	0.07	LM	0	3.68	1.62	LM
	Phosphorus	-18	3.15	3.95	LM	-2	3.18	2.71	LM

Table D2

A summary of the ranges of the predictor variables from each dataset, and the number of novel environments for each predictor variable per dataset. Lapalala has 180 sites and Welgevonden has 65 sites.

Predictor variable	Lapalala						Welgevonden						# W sites < L Min[1]	# W sites > L Max[2]	# L sites < W Min[3]	# L sites > W Max[4]
	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Min	1st Quartile	Median	Mean	3rd Quartile	Max				
Elevation	921	1117	1165	1158	1197	1292	1134	1250	1332	1338	1401	1661	0	40	66	0
Slope	0.000	1.949	2.923	3.778	4.480	19.687	0.972	3.070	6.198	7.166	10.042	30.121	0	1	8	0
Curvature	-0.035	-0.002	0.000	0.000	0.005	0.014	-0.028	-0.007	0.002	0.001	0.012	0.032	0	11	2	0
TWI	0.197	3.685	10.288	7.873	10.981	14.366	0.398	3.576	10.291	7.217	10.291	15.699	0	1	1	0
PDIR	0.160	0.979	1.049	1.018	1.093	1.180	0.813	0.976	1.021	1.015	1.066	1.172	0	0	12	2
Rock cover	0	0	5	19	30	90	0	0	8	19	35	85	0	0	0	1
Woody debris cover	0	1	2	4	5	40	0	2	5	10	15	30	0	0	0	1
Bare soil cover	0	10	20	24	33	80	1	2	8	13	18	40	0	0	3	32
Leaf litter cover	0	5	10	18	25	80	1	10	20	20	30	50	0	0	7	12
Soil pH	4.02	4.74	5.13	5.22	5.62	6.91	4.41	4.76	4.99	5.16	5.48	6.44	0	0	11	4
Clay	6.12	9.95	10.20	11.42	12.24	24.49	11.00	15.00	15.00	15.77	17.00	33.00	0	1	91	0
Carbon	0.01	0.38	0.73	1.03	1.58	4.47	0.32	0.88	1.36	1.59	1.97	4.94	0	1	35	0
Potassium	2.08	4.82	6.80	8.67	9.34	51.15	22.54	57.69	72.23	78.47	98.89	153.69	0	53	173	0
Calcium	12.78	24.02	35.89	66.60	80.69	556.00	23.70	90.32	161.12	233.23	230.32	1251.02	0	5	45	0
Sodium	1.75	2.40	2.65	2.71	2.98	4.97	6.53	9.89	11.65	11.60	12.99	22.98	0	65	180	0
Phosphorus	0.15	1.06	1.63	1.88	2.57	7.19	4.43	7.80	10.55	12.25	14.16	36.23	0	52	175	0

[1] The number of sites in the W dataset that have a value for the specified predictor variable less than the minimum value for the same predictor variable in the L dataset. [2] The number of sites in the W dataset that have a value for the specified predictor variable greater than the maximum value for the same predictor variable in the L dataset. [3] The number of sites in the L dataset that have a value for the specified predictor variable less than the minimum value for the same predictor variable in the W dataset. [4] The number of sites in the L dataset that have a value for the specified predictor variable greater than the maximum value for the same predictor variable in the W dataset.

Table D3

A summary of the predictor variables that were included in the various models. + = a positive or - = a negative relationship with the response variable in that model. Where a quadratic term was included for the variable, the first sign represents the linear term and the second the quadratic term. The training dataset (L: Lapalala, W: Welgevonden) is the dataset that was used to build the model. The model type column refers to the three modelling approaches (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model).

Model ID	Response variable	Training dataset	Model type	Number of terms	Adjusted deviance explained (%)	Elevation	Slope	Curvature	TWI	PDIR	Rock cover	Woody debris cover	Bare soil cover	Leaf litter cover	Soil pH	Clay	Carbon	Potassium	Calcium	Sodium	Phosphorus
1	Total species	L	1	17	23.3	-	-	+	+	+	+	-	-	+	+	+	+	-	-	+	+/-
2		L	2	9	25.8		-				+			+	+	+	+	-			+/-
3		L	3	6	19.0		-				+		-	+	+		+				
4		W	1	20	40.3	-/+	+	+	-	+	+	+/-	-	+	+/-	-	-/-	+	+	+	-
5		W	2	12	46.3		+	+	-			+/-			+/-		+/-	+	+		-
6		W	3	10	43.0		+	+	-			+/-	-		+/-		-		+		
7	Grass species	L	1	16	10.3	+	-	-	-	-	+	-	-	+	+	+	-	+	+	+	+
8		L	2	4	13.1				-			-			+					+	
9		L	3	3	11.3							-			+				+		
10		W	1	18	41.7	-	-	-	-	-	+	-	-	-	+/-	+	-	+	+/-	-	-
11		W	2	6	44.0										+/-		-	+		-	-
12		W	3	7	34.5				-				-		+/-		-		+/-		
13	Herb species	L	1	17	10.0	+	-	+	+	+	+	-	+	+	+	+	-	-	+	+	+/-
14		L	2	9	12.9		-			+	+	-			+	+		-			+/-
15		L	3	5	4.1		-			+	+	-			+						
16		W	1	18	38.9	-	+	+	-	-	+	+	+	+	+/-	-	+/-	+	+	+	-
17		W	2	8	41.6			+	-			+			+/-		+/-	+			
18		W	3	9	37.6			+	-		+	+			+/-		+/-		+		

Table D3. Continued.

Model ID	Richness variable	Training dataset	Model type	Number of terms	Adjusted deviance explained (%)	Elevation	Slope	Curvature	TWI	PDIR	Rock cover	Woody debris cover	Bare soil cover	Leaf litter cover	Soil pH	Clay	Carbon	Potassium	Calcium	Sodium	Phosphorus
19	Woody species	L	1	19	45.2	+	-	-	-	+	+	+	-	+	+	+	+/-	-	+/-	-	+/-
20		L	2	8	47.5					+	+			+	+		+/-				+/-
21		L	3	6	45.5					+	+			+	+		+/-				
22		W	1	21	60.1	+	+	+	-	+/-	+	+/-	-	+	-	-	-/+	-	+	+/-	-/-
23		W	2	7	64.4			+			+	+/-			-				+		-
24		W	3	7	58.9	+	+					+/-		+	-				+		
25	Family	L	1	18	33.4	+	-	+	+	+	+	-	-	+	+	+	+	-/+	+	+	+/-
26		L	2	7	34.5					+	+				+		+	-			+/-
27		L	3	4	28.9					+	+			+			+				
28		W	1	19	37.5	+	+	+	-	+/-	+	+/-	-	+	-	-	+/-	-	+	+	+
29		W	2	6	39.9		+	+	-			+/-									+
30		W	3	7	42.0		+	+	-			+/-					+/-				
31	Genus	L	1	16	20.1	+	-	+	-	+	+	-	-	+	+	+	+	-	-	+	-
32		L	2	10	22.6	+	-				+	-	-	+	+		+	-		+	
33		L	3	7	21.3		-				+	-	-	+	+		+				
34		W	1	19	44.0	-	+	+	-	+	+	+/-	-	+	+/-	-	-/-	+	+	-	-
35		W	2	12	49.5		+	+	-			+/-	-		+/-		-	+	+		-
36		W	3	10	46.7		+	+	-			+/-	-		+/-		-		+		

Table D4

A summary of the transferability results using the unclamped and the clamped datasets. The training dataset (L: Lapalala, W: Welgevonden) was used to build the model with which the testing dataset was then used to predict richness. The model type column refers to the three modelling approaches (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model). The AIC and adjusted deviance explained (adj. DE) values are for the GLMs used to predict richness and the slope estimate, p-value and R^2 values are for the linear models used to test transferability. The number of terms in each model counts quadratic terms as a second term thus there can be more than 16 terms in a model.

									Unclamped dataset			Clamped dataset		
Model ID	Richness variable	Training dataset	Model type	Number of sites	Number of terms	AIC	Adj. DE (%)	Testing dataset	Slope estimate	P-value	R^2	Slope estimate	P-value	R^2
1	Total species	L	1	172	18	1410	23.3	W	-0.028	0.776	0.001	0.023	0.527	0.007
2	Total species	L	2	172	10	1400	25.8	W	-0.035	0.670	0.003	0.022	0.561	0.006
3	Total species	L	3	172	7	1441	19.0	W	0.027	0.656	0.004	0.047	0.427	0.011
4	Total species	W	1	60	21	448	40.3	L	-0.080	0.156	0.012	-0.082	0.108	0.015
5	Total species	W	2	60	13	437	46.3	L	-0.061	0.094	0.016	-0.061	0.094	0.016
6	Total species	W	3	60	11	439	43.0	L	-0.065	0.150	0.012	-0.072	0.094	0.016
7	Grass species	L	1	170	17	845	10.3	W	0.055	0.922	0.000	0.136	0.052	0.065
8	Grass species	L	2	170	5	827	13.1	W	-0.392	0.370	0.014	0.073	0.072	0.056
9	Grass species	L	3	170	4	829	11.3	W	-0.030	0.688	0.003	0.049	0.301	0.019
10	Grass species	W	1	59	19	335	41.7	L	-0.173	0.059	0.021	-0.174	0.028	0.028
11	Grass species	W	2	59	7	317	44.0	L	-0.182	0.003	0.051	-0.168	0.002	0.057
12	Grass species	W	3	59	8	324	34.5	L	-0.110	0.131	0.014	-0.115	0.087	0.017
13	Herb species	L	1	177	18	1317	10.0	W	-0.029	0.308	0.019	-0.033	0.287	0.021
14	Herb species	L	2	177	10	1309	12.9	W	-0.025	0.243	0.025	-0.032	0.199	0.030
15	Herb species	L	3	177	6	1358	4.1	W	-0.039	0.296	0.020	-0.030	0.379	0.014
16	Herb species	W	1	57	19	367	38.9	L	-0.019	0.739	0.001	0.011	0.838	0.000
17	Herb species	W	2	57	9	357	41.6	L	-0.025	0.450	0.003	-0.012	0.722	0.001
18	Herb species	W	3	57	10	362	37.6	L	0.022	0.610	0.002	0.031	0.461	0.003
19	Woody species	L	1	170	20	1021	45.2	W	-0.018	0.597	0.005	0.074	0.012	0.101
20	Woody species	L	2	170	9	1005	47.5	W	-0.012	0.885	0.000	0.094	0.014	0.098
21	Woody species	L	3	170	7	1011	45.5	W	0.273	0.000	0.348	0.297	0.000	0.409
22	Woody species	W	1	62	22	371	60.1	L	0.047	0.236	0.008	0.052	0.282	0.007
23	Woody species	W	2	62	8	354	64.4	L	0.097	0.048	0.023	0.102	0.023	0.030
24	Woody species	W	3	62	8	365	58.9	L	0.107	0.001	0.062	0.089	0.004	0.047

Table D4. Continued.

Model ID	Richness variable	Training dataset	Model type	Number of sites	Number of terms	AIC	Adj. DE (%)	Testing dataset	Unclamped dataset Slope estimate	Unclamped dataset P-value	Unclamped dataset R^2	Clamped dataset Slope estimate	Clamped dataset P-value	Clamped dataset R^2
25	Family	L	1	172	19	1026	33.4	W	-0.466	0.023	0.089	0.137	0.011	0.109
26		L	2	172	8	1013	34.5	W	-0.174	0.005	0.134	0.157	0.001	0.174
27		L	3	172	5	1024	28.9	W	0.308	0.003	0.148	0.304	0.001	0.170
28		W	1	58	20	355	37.5	L	0.058	0.122	0.014	0.070	0.054	0.022
29		W	2	58	7	338	39.9	L	0.064	0.019	0.032	0.051	0.073	0.019
30		W	3	58	8	338	42.0	L	0.272	0.000	0.234	0.251	0.000	0.211
31	Genus	L	1	175	17	1373	20.1	W	-0.124	0.171	0.031	-0.035	0.451	0.010
32		L	2	175	11	1363	22.6	W	-0.137	0.123	0.039	-0.038	0.462	0.009
33		L	3	175	8	1371	21.3	W	-0.052	0.345	0.015	-0.035	0.513	0.007
34		W	1	62	20	447	44.0	L	-0.080	0.124	0.014	-0.104	0.039	0.024
35		W	2	62	13	436	49.5	L	-0.115	0.012	0.036	-0.145	0.002	0.057
36		W	3	62	11	438	46.7	L	-0.069	0.131	0.013	-0.097	0.026	0.028

Appendix E

Generalised Boosted Models (GBMs) Results

Transferability of the richness models was consistently poor for both the GLMs and GBMs. Here results are reported for the GBMs (see the text for the results of the GLMs, which were largely similar).

1.1. Model building

The performance of the GBMs varied, with Pearson correlation (10-fold cross validation) values ranging from 0.70 to 0.04. Woody species richness and family richness were the two response variables for which the GBMs performed the best (Pearson correlation values of 0.64 and 0.60 averaged across the three modelling approaches, respectively). In contrast, grass and herb species richness were the two response variables for which GBMs performed the worst (Pearson correlation values of 0.38 and 0.27 averaged across the three modelling approaches, respectively). Minimum Adequate Models performed the best, followed by the Shared Environments Models. The predictor variables which were in the greatest number of GBMs were soil pH and soil carbon, in agreement with the GLMs results.

1.2. Transferability

All 72 GBMs (36 clamped and 36 unclamped) displayed poor transferability regardless of the model type and whether the datasets were clamped or unclamped (**Table E1** and **Fig. E1**). The slope estimate values were not close to one (the highest value being 0.38) and the R^2 values were also low (the highest being 0.44). The majority of the GBM outputs (49 out of 72) were statistically significant (p-value < 0.05), meaning the correlation between the predicted and observed richness values performed better than random. Similarly to GLMs, the measures of richness which displayed the best transferability for the GBMs, albeit weak, were woody species richness and family richness as they displayed the slope estimates closest to 1 and with the highest R^2 values.

Clamping the testing dataset to predict richness improved the transferability for 20 out of 36 of the approaches (i.e. displayed a slope estimate closer to 1). Of the three modelling approaches, the Shared Environments Models displayed the best transferability for the GBMs (consistent with results from the GLMs).

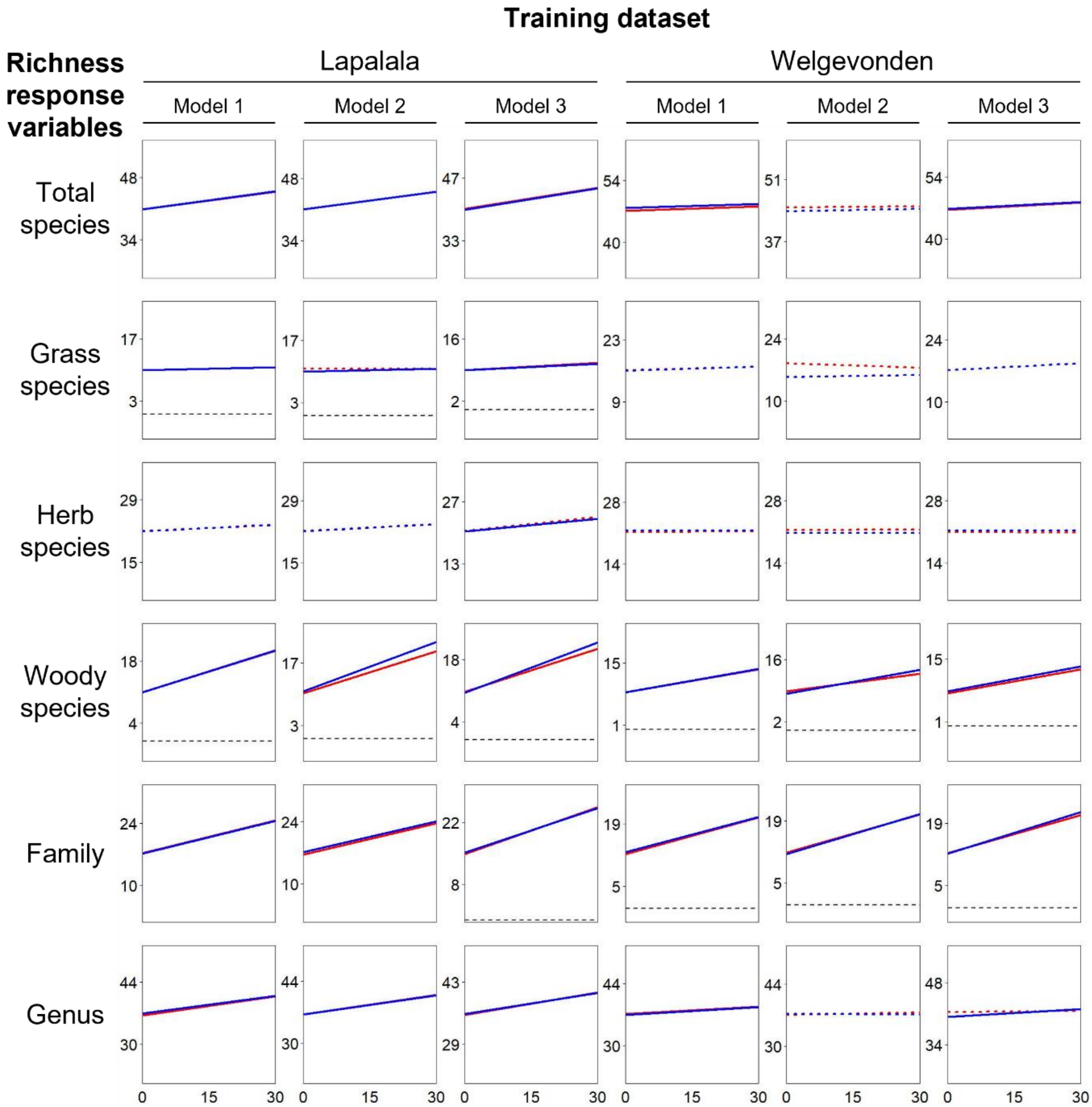

Fig. E1. Comparing the predicted and observed richness values for the three different models (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model). The training dataset is the dataset used to build the models. The red lines are the unclamped dataset's predicted richness against the observed richness, and the blue lines are the clamped dataset's predicted richness against observed richness (solid line: the output was statistically significant [< 0.05], dashed line: the output was not statistically significant). The dashed black horizontal line is at y = 0. All the y-axes and x-axes have the same range (30 units).

Table E1

A summary of the model building and transferability results of the Generalised Boosted Models (GBMs). The training dataset (L: Lapalala, W: Welgevonden) was used to build the model with which the testing dataset was then used to predict richness. The model type column refers to the three modelling approaches (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model). The clamped column refers to if the testing dataset was clamped (Y) or unclamped (N). The Pearson correlation (cross validation) column is for the GBMs used to predict richness and the slope estimate, p-value and R^2 values are for the linear models used to test transferability.

Model ID	Richness variable	Training dataset	Model type	Number of sites	Number of predictors	Clamped?	Testing dataset	Number of trees	Pearson correlation	Slope estimate	P-value	R^2
1	Species Richness	L	1	60	16	N	W	3650	0.56	0.132	0.002	0.158
2		L	1	60	16	Y	W	3450	0.55	0.136	0.002	0.157
3		L	2	60	11	N	W	2850	0.54	0.132	0.002	0.156
4		L	2	60	11	Y	W	2900	0.57	0.132	0.002	0.157
5		L	3	60	11	N	W	2850	0.54	0.157	0.001	0.162
6		L	3	60	8	Y	W	2600	0.52	0.162	0.001	0.162
7		W	1	172	16	N	L	1600	0.33	0.033	0.027	0.028
8		W	1	172	16	Y	L	1350	0.31	0.028	0.037	0.025
9		W	2	172	3	N	L	3300	0.58	0.007	0.781	0.000
10		W	2	172	2	Y	L	2950	0.34	0.020	0.368	0.005
11		W	3	172	8	N	L	2400	0.43	0.057	0.009	0.039
12		W	3	172	7	Y	L	3550	0.48	0.051	0.049	0.023
13	Grass Richness	L	1	60	16	N	W	1200	0.26	0.023	0.013	0.103
14		L	1	60	16	Y	W	1200	0.27	0.024	0.006	0.124
15		L	2	60	4	N	W	1050	0.30	0.003	0.755	0.002
16		L	2	60	13	Y	W	1100	0.28	0.021	0.017	0.094
17		L	3	60	12	N	W	1500	0.21	0.057	0.012	0.104
18		L	3	60	12	Y	W	1150	0.28	0.048	0.010	0.109
19		W	1	172	16	N	L	4200	0.47	0.026	0.423	0.004
20		W	1	172	16	Y	L	4250	0.52	0.027	0.406	0.004
21		W	2	172	2	N	L	3500	0.43	-0.030	0.545	0.002
22		W	2	172	6	Y	L	4950	0.61	0.014	0.749	0.001
23		W	3	172	4	N	L	4250	0.37	0.053	0.346	0.005
24		W	3	172	4	Y	L	4100	0.54	0.053	0.343	0.005

Table E1. Continued.

Model ID	Richness variable	Training dataset	Model type	Number of sites	Number of predictors	Clamped?	Testing dataset	Number of trees	Pearson correlation	Slope estimate	P-value	R^2
25	Herb Richness	L	1	60	16	N	W	1650	0.24	0.046	0.146	0.036
26		L	1	60	16	Y	W	2000	0.31	0.050	0.166	0.033
27		L	2	60	8	N	W	2150	0.38	0.052	0.098	0.047
28		L	2	60	8	Y	W	1500	0.33	0.055	0.110	0.044
29		L	3	60	6	N	W	3350	0.34	0.108	0.074	0.054
30		L	3	60	7	Y	W	2000	0.34	0.096	0.042	0.069
31		W	1	172	16	N	L	2100	0.04	0.003	0.570	0.002
32		W	1	172	16	Y	L	3000	0.07	0.002	0.563	0.002
33		W	2	172	7	N	L	1950	0.28	0.003	0.523	0.002
34		W	2	172	6	Y	L	8800	0.23	0.000	0.987	0.000
35		W	3	172	3	N	L	7600	0.26	-0.002	0.901	0.000
36		W	3	172	4	Y	L	9200	0.36	-0.001	0.939	0.000
37	Woody Richness	L	1	60	16	N	W	4700	0.69	0.312	0.000	0.438
38		L	1	60	16	Y	W	3700	0.68	0.314	0.000	0.434
39		L	2	60	8	N	W	6050	0.70	0.316	0.000	0.400
40		L	2	60	3	Y	W	3150	0.69	0.369	0.000	0.399
41		L	3	60	4	N	W	3450	0.69	0.326	0.000	0.413
42		L	3	60	3	Y	W	3450	0.67	0.376	0.000	0.400
43		W	1	172	16	N	L	10000	0.66	0.173	0.000	0.178
44		W	1	172	16	Y	L	9950	0.57	0.179	0.000	0.190
45		W	2	172	10	N	L	5000	0.51	0.131	0.000	0.132
46		W	2	172	15	Y	L	10000	0.54	0.179	0.000	0.189
47		W	3	172	11	N	L	8700	0.58	0.179	0.000	0.192
48		W	3	172	11	Y	L	4050	0.66	0.185	0.000	0.246

Table E1. Continued.

Model ID	Richness variable	Training dataset	Model type	Number of sites	Number of predictors	Clamped?	Testing dataset	Number of trees	Pearson correlation	Slope estimate	P-value	R^2
49	Family Richness	L	1	60	16	N	W	3550	0.58	0.243	0.000	0.378
50		L	1	60	16	Y	W	3400	0.60	0.248	0.000	0.372
51		L	2	60	4	N	W	3600	0.60	0.236	0.000	0.307
52		L	2	60	4	Y	W	3150	0.61	0.230	0.000	0.325
53		L	3	60	5	N	W	5600	0.61	0.352	0.000	0.304
54		L	3	60	2	Y	W	2800	0.60	0.333	0.000	0.331
55		W	1	172	16	N	L	5050	0.61	0.277	0.000	0.263
56		W	1	172	16	Y	L	4150	0.57	0.266	0.000	0.263
57		W	2	172	13	N	L	6200	0.64	0.288	0.000	0.252
58		W	2	172	9	Y	L	6200	0.55	0.300	0.000	0.244
59		W	3	172	9	N	L	3800	0.68	0.286	0.000	0.248
60		W	3	172	7	Y	L	4100	0.55	0.311	0.000	0.253
61	Genus Richness	L	1	60	16	N	W	3800	0.55	0.141	0.001	0.169
62		L	1	60	16	Y	W	3050	0.53	0.132	0.001	0.175
63		L	2	60	14	N	W	3250	0.54	0.144	0.001	0.163
64		L	2	60	14	Y	W	4000	0.58	0.145	0.001	0.169
65		L	3	60	10	N	W	3150	0.56	0.168	0.001	0.179
66		L	3	60	11	Y	W	2750	0.53	0.156	0.001	0.183
67		W	1	172	16	N	L	3000	0.36	0.055	0.006	0.044
68		W	1	172	16	Y	L	3100	0.32	0.057	0.006	0.044
69		W	2	172	2	N	L	3250	0.45	0.016	0.478	0.003
70		W	2	172	4	Y	L	4950	0.46	-0.003	0.901	0.000
71		W	3	172	3	N	L	3500	0.45	0.007	0.819	0.000
72		W	3	172	12	Y	L	1450	0.16	0.060	0.001	0.058