

Fine-tuned ecological niche models unveil climatic suitability and association with vegetation groups for selected *Chaetocnema* species in South Africa (Coleoptera: Chrysomelidae)

Francesco Cerasoli^{1*}, Paola D'Alessandro¹ & Maurizio Biondi¹

¹ Department of Life, Health and Environmental Sciences - Environmental Sciences Section, University of L'Aquila, L'Aquila, Italy

*Correspondence: Francesco Cerasoli, University of L'Aquila, Via Vetoio, 67100 Coppito, L'Aquila, Italy; email: francesco.cerasoli@univaq.it

Supplementary Materials

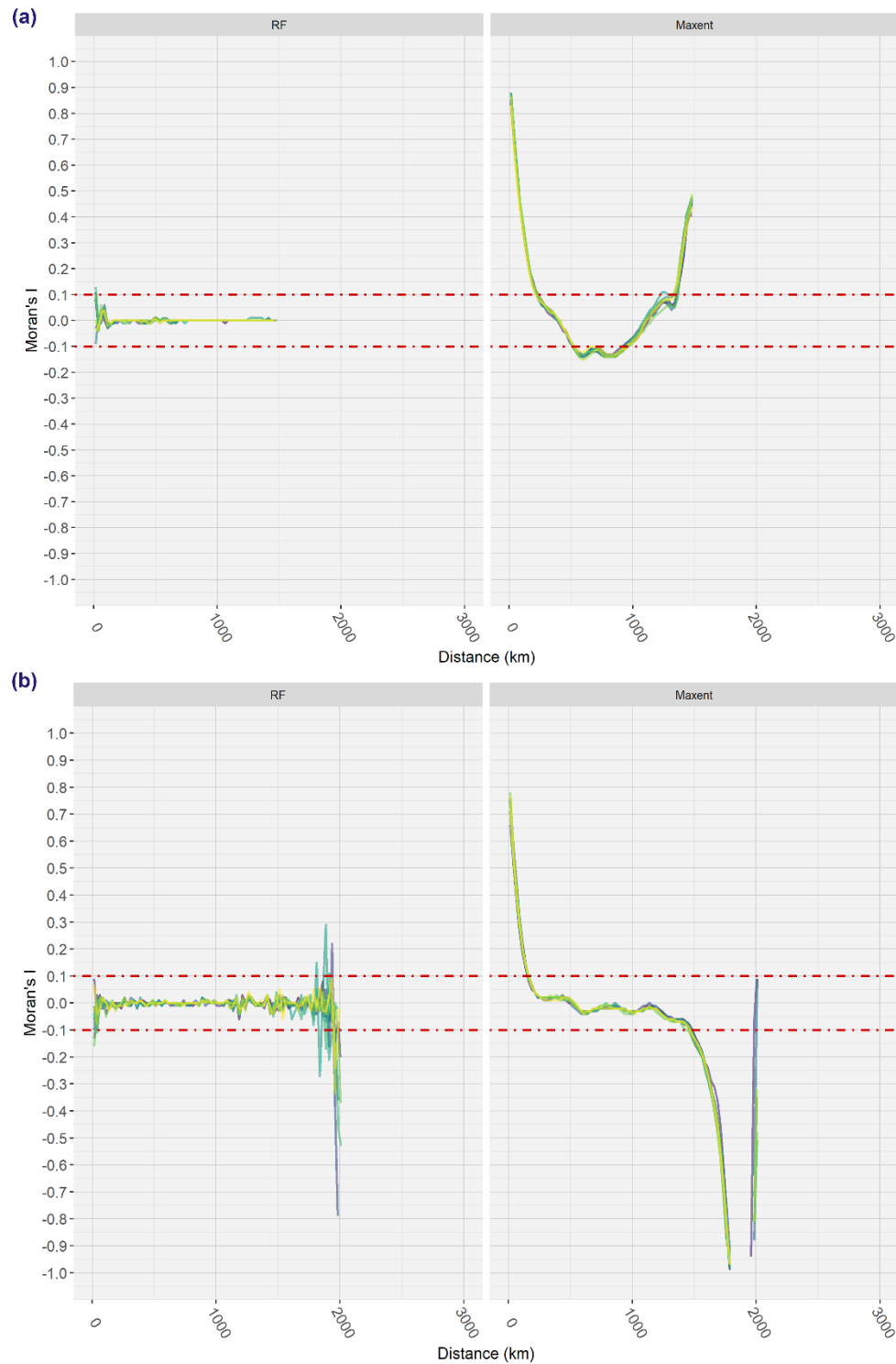
Table S1 Target species' occurrence coordinates. Coordinates (WGS84 geographic coordinate system) of occurrence records used to fit the Ecological Niche Models for each target *Chaetocnema* species.

Species	Long_WGS84	Lat_WGS84
<i>C. brincki</i>	29,07	-29,28
	29,43	-29,65
	29,45	-29,95
	29,5	-29,78
	29,98	-29,37
	30,48	-25,17
<i>C. danielssoni</i>	18,41	-33,96
	18,42	-33,97
	19,02	-32,67
	19,08	-31,37
	19,23	-34,42
	19,37	-33,62
	19,62	-34,05
	23,05	-34,03
	23,62	-33,95
	24,28	-33,95
<i>C. darwini</i>	18,2	-31,67
	18,35	-34,05
	18,37	-33,37

	18,38	-34,27
	18,38	-34,08
	18,42	-33,97
	18,48	-34,35
	18,65	-33,12
	18,85	-33,92
	19,02	-32,67
	19,08	-33,6
	19,23	-34,42
	19,3	-33,35
	19,37	-33,62
	19,42	-34,43
	19,67	-33,37
	20,05	-34,8
	20,05	-34,62
	22,47	-33,97
	22,55	-33,98
	22,58	-33,95
	22,65	-33,98
	23,05	-34,03
	23,15	-33,93
	23,37	-34,08
	23,37	-34,05
	23,39	-34,03
	23,5	-33,95
	23,55	-33,97
	23,63	-33,95
	23,92	-33,98
	24,28	-33,95
	26,53	-33,32
	27,22	-31,28
	28,47	-28,53
	29,43	-25,58
	29,5	-29,78
<i>C. gahani</i>	26,67	-32,5
	27,2	-25,67
	27,23	-25,65
	27,38	-24,62
	27,75	-24,57
	27,75	-23,97
	28,47	-28,53
	29,15	-24,05
	29,77	-28,88
	29,87	-26,14
	30,27	-25,6

	30,92	-29,87
<i>C. natalensis</i>	18,29	-33,57
	18,43	-33,92
	18,48	-34,05
	19,62	-34,05
	20,82	-34,42
	22,1	-34,13
	22,2	-33,42
	22,4	-33,65
	23,37	-34,08
	26,47	-33,37
	26,97	-31,9
	27,52	-27,22
	28,3	-25,68
	28,38	-32,67
	28,63	-24,22
	29,4	-29,05
	29,5	-29,78
	29,52	-31,62
	29,83	-23,02
	29,95	-23,94
	30,48	-25,17
	30,92	-29,87
	30,93	-29,83
	31	-29,87
	31,71	-29
	32,17	-27,28

Figure S1 Correlograms from preliminary models. Correlograms representing variation in Moran index (I) at increasing inter-point distance (25 km pace), computed on residuals from predictions on training data of the 10 preliminary Random Forest (RF) and 10 preliminary Maxent models fitted to choose block size for: (a) *Chaetocnema darwini*; (b) *C. natalensis*.



Each colour represents a different preliminary model; red dash-dotted lines represent the interval $-0.1 - 0.1$ in Moran's I.

Figure S2 Checkerboard cross-validation blocks. Checkerboard spatial blocking structure splitting between training and test folds the presence (purple triangles) and background (grey crosses) data used to fit the RF models for: (a) *Chaetocnema darwini*; (b) *C. natalensis*.

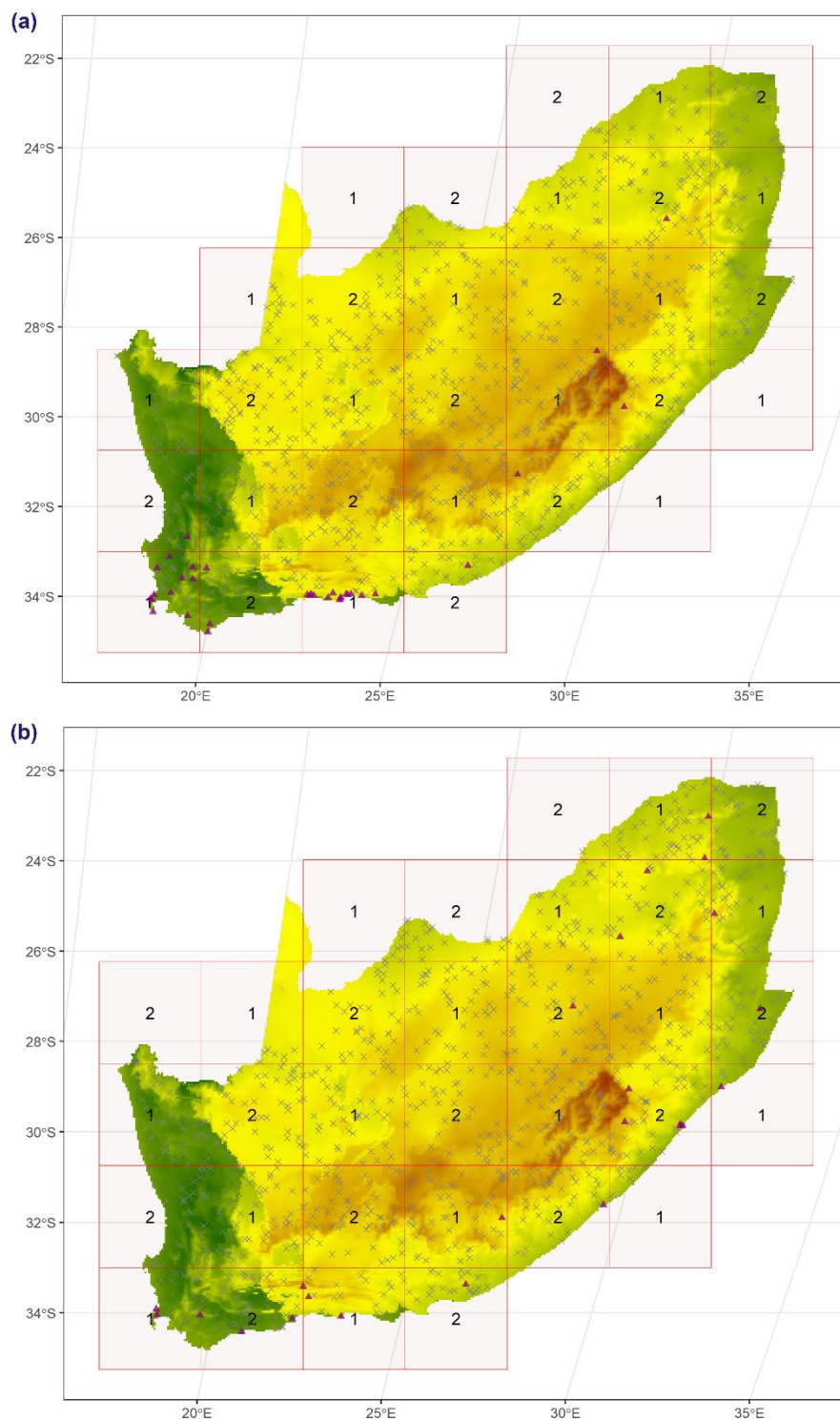
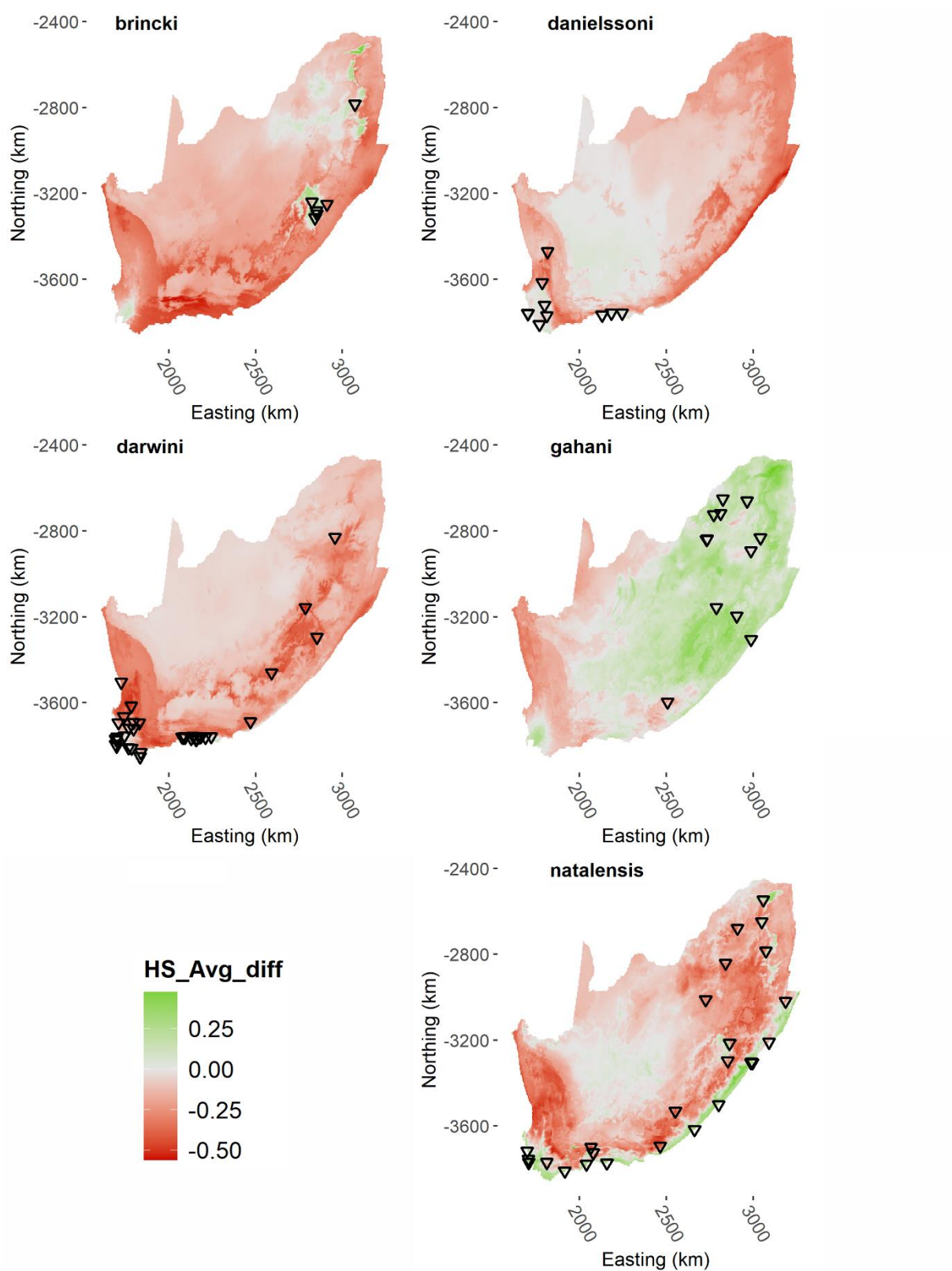


Table S2 Parameterization and discrimination performance of tuned models. Tuned parameters for both Maxent and Random Forests are reported for each *Chaetocnema* species × cross-validation split combination, along with AUC values attained by the tuned models on the test fold (AUCtest) and with the difference between AUC attained by the same models on the training data and AUCtest (AUCdiff).

Species	CV split	Maxent FC_RM	Maxent AUCtest	Maxent AUCdiff	RF Trees	RF Node size	RF AUCtest	RF AUCdiff
<i>brincki</i>	LOO_1	LQH_1	0,97	0	1000	5	0,62	0,37
	LOO_2	LQ_1	0,98	0,01	2000	1	0,88	0,12
	LOO_3	L_1.5	0,95	0	2000	1	0,75	0,24
	LOO_4	L_3	0,96	0	2000	5	0,8	0,19
	LOO_5	L_1.5	0,88	0,08	500	1	0,91	0,09
	LOO_6	L_1.5	0,88	0,08	2000	5	0,75	0,25
<i>danielssoni</i>	LOO_1	LQH_1	0,98	0,01	1000	1	1	-0,01
	LOO_2	LQH_1	0,98	0,01	500	1	1	-0,01
	LOO_3	L_3	0,96	0	2000	5	0,61	0,38
	LOO_4	L_3	0,91	0,05	500	5	0,86	0,13
	LOO_5	LQH_1	0,97	0	500	1	0,96	0,03
	LOO_6	LQH_3	0,97	0	500	1	0,78	0,21
	LOO_7	LQH_3	0,94	0,04	500	5	0,94	0,05
	LOO_8	L_1	0,94	0,03	1000	1	0,9	0,09
	LOO_9	L_1	0,94	0,03	1000	1	0,86	0,13
	LOO_10	L_1	0,93	0,03	2000	3	0,78	0,21
<i>darwini</i>	Checkerboard_1	LQHPT_3	0,91	0,03	2000	5	0,97	0
	Checkerboard_2	L_1.5	0,9	0,09	1000	5	0,87	0,13
<i>gahani</i>	LOO_1	LQH_2	0,51	0,3	500	3	0,7	0,28
	LOO_2	LQH_2.5	0,81	0,05	500	1	1	0
	LOO_3	LQH_2.5	0,81	0,08	500	3	0,67	0,31
	LOO_4	LQH_1	0,89	0	1000	5	0,94	0,02
	LOO_5	LQH_2	0,86	0,01	1000	1	0,67	0,32
	LOO_6	LQH_1	0,77	0,12	1000	3	0,6	0,38
	LOO_7	LQ_3	0,81	0,06	1000	1	1	-0,02
	LOO_8	LQH_1.5	0,88	0,01	500	3	0,89	0,11
	LOO_9	LQH_1.5	0,87	0,05	2000	5	0,67	0,31
	LOO_10	LQH_1	0,88	0,01	500	3	0,94	0,04
	LOO_11	LQH_1	0,88	0,01	2000	1	0,71	0,26
	LOO_12	LQH_1	0,88	0,01	1000	1	1	-0,01
<i>natalensis</i>	Checkerboard_1	L_3	0,89	0	500	5	0,86	0,09
	Checkerboard_2	L_3	0,84	0,04	500	5	0,79	0,19

FC = feature class; RM = regularization multiplier; Trees = number of trees in the forest; Node size = number of observations in each terminal node of the trees.

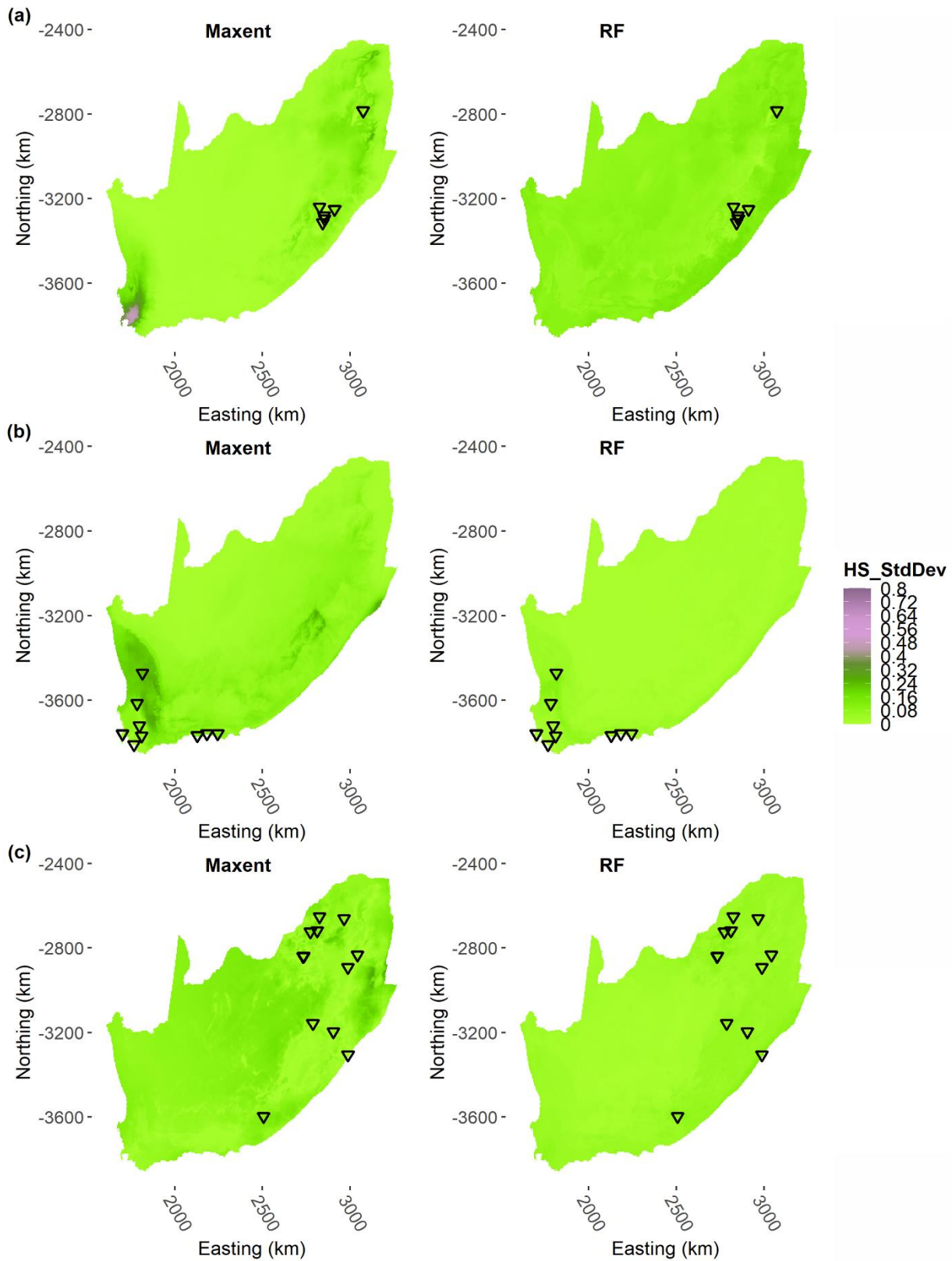
Figure S3 Between-algorithm suitability difference. Difference between Maxent-derived and RF-derived weighted average habitat suitability (HS_Avg_diff.) computed across the train-test CV splits for each target *Chaetocnema* species.



Reversed black-contoured triangles indicate the species' occurrence localities.

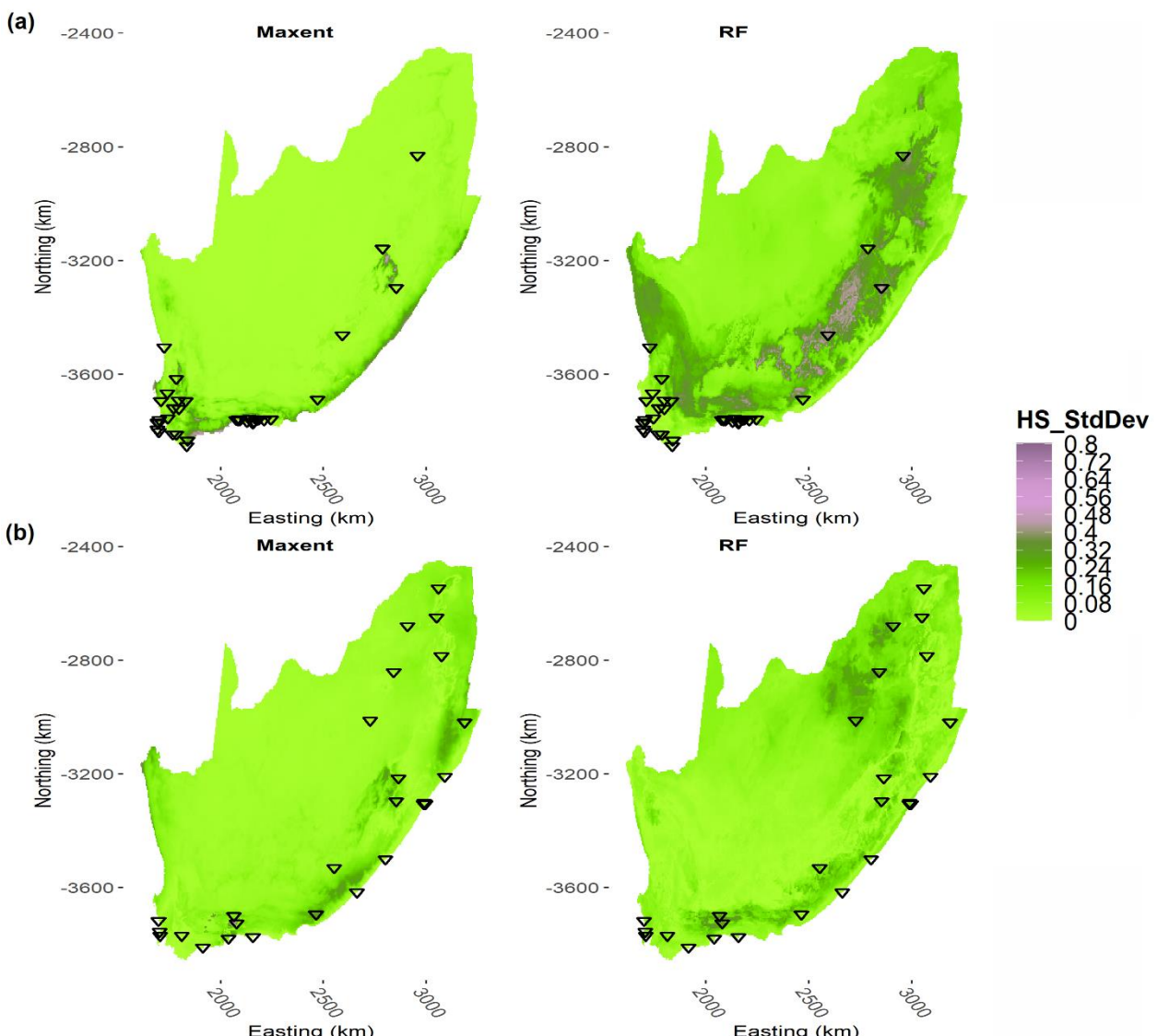
Figure S4 Suitability standard deviation for *Chaetocnema brincki*, *C. danielssoni* and *C. gahani*.

Weighted average standard deviation in predicted habitat suitability (HS) among the train-test CV splits for: (a) *Chaetocnema brincki*; (b) *C. danielssoni*; (c) *C. gahani*.



Reversed black-contoured triangles indicate the species' occurrence localities.

Figure S5 Suitability standard deviation for *Chaetocnema darwini* and *C. natalensis*. Weighted average standard deviation in predicted habitat suitability (HS) among the train-test CV splits for: (a) *Chaetocnema darwini*; (b) *C. natalensis*.



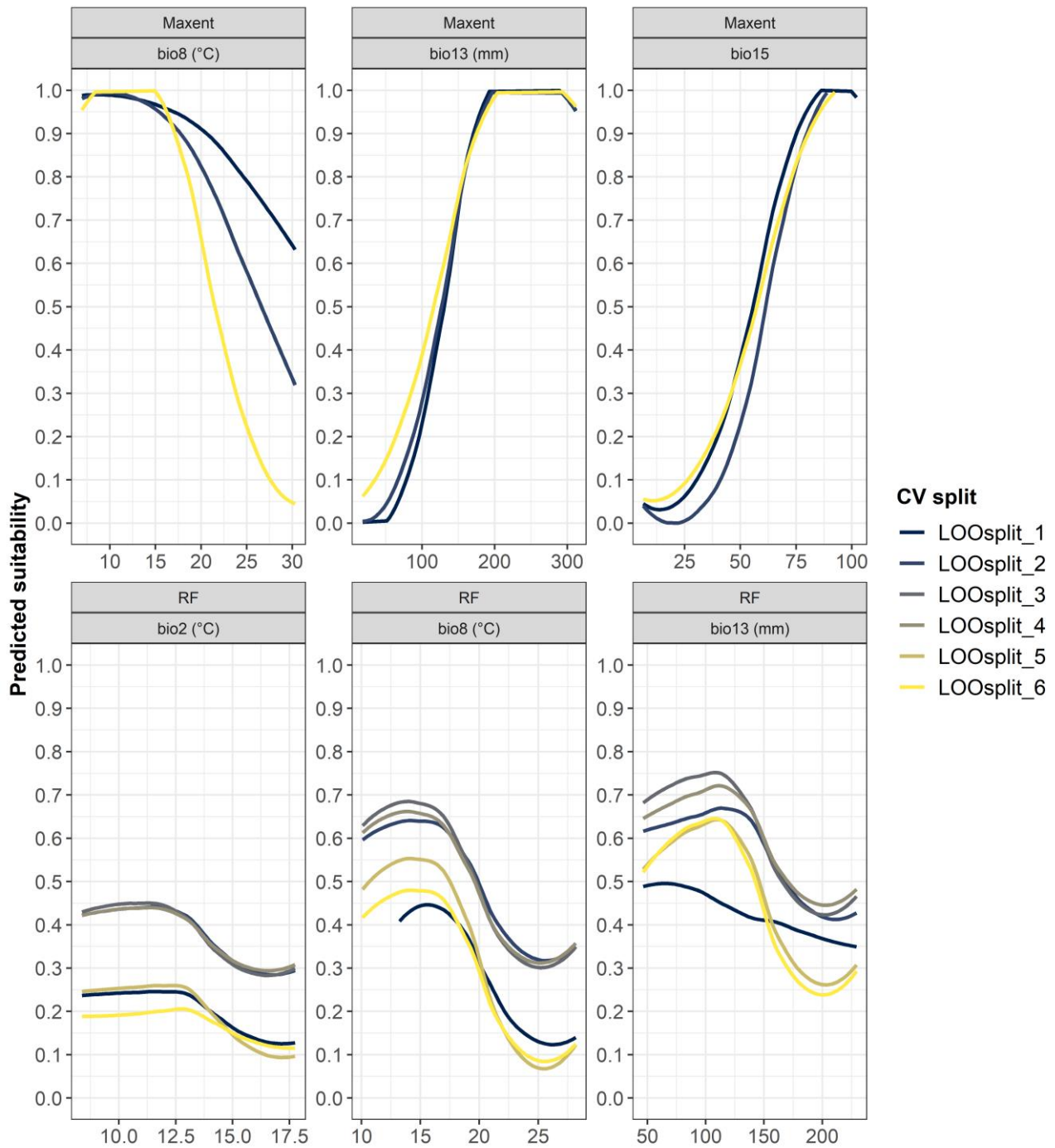
Reversed black-contoured triangles indicate the species' occurrence localities.

Table S3 Variables importance scores. Weighted average (Wgt_Avg) and weighted standard deviation (Wgt_StdDev) of standardized importance scores computed for the selected modelling variables within the tuned Maxent and RF models.

Species	Variable	Maxent Wgt_Avg	Maxent Wgt_StdDev	RF Wgt_Avg	RF Wgt_StdDev
<i>C. brincki</i>	bio2	0,01	0,02	15,49	2,97
	bio3	0,11	0,26	7,9	1,91
	bio8	26,05	19,25	23,88	0,72
	bio9	14,02	21,36	12,18	2,58
	bio13	31,17	6,22	19,84	3,17
	bio15	28,65	3,63	10,1	0,98
	bio19	0	0	10,57	1,49
<i>C. danielssoni</i>	bio2	0,15	0,11	16,85	2,03
	bio3	0	0	4,16	0,91
	bio8	21,53	31,3	20,47	1,81
	bio9	0,02	0,02	12,33	1,25
	bio13	0,3	0,28	10,97	1,61
	bio15	0,74	0,66	4,12	0,8
	bio19	77,27	30,73	31,08	0,59
<i>C. darwini</i>	bio2	40,79	56,94	18,93	6,78
	bio3	0	0	5,94	4,59
	bio8	40,66	38,92	28,05	7,55
	bio9	0,53	0,25	8,93	2,89
	bio13	9,89	14,06	7,8	2,54
	bio15	0,96	1,24	7,21	1,48
	bio19	7,17	4,96	23,09	6,42
<i>C. gahani</i>	bio2	37,19	18,36	20,83	1,59
	bio3	3,31	2,83	12,07	1,22
	bio8	0	0	11,62	0,7
	bio9	0,45	0,45	12,28	0,9
	bio13	49,41	26,47	20,12	1,57
	bio15	8,33	4,66	13,18	1,13
	bio19	1,32	1,33	9,9	0,94
<i>C. natalensis</i>	bio2	91,06	11,25	28,18	6,01
	bio3	2,87	3,8	9,79	1,34
	bio8	0,17	0,24	9,82	0,92
	bio9	1,73	2,37	11,58	0,85
	bio13	3,59	4,12	13,62	5,86
	bio15	0,54	0,74	10,16	2,12
	bio19	0,04	0,03	16,86	1,41

Scores of the top three variables in terms of Wgt_Avg relative importance are in red.

Figure S6 Response curves for *Chaetocnema brincki*. For *Chaetocnema brincki*, marginal response curves obtained from the tuned Maxent and RF models fitted for each train-test leave-one-out (LOO) split, for the top three variables in terms of weighted average standardized importance scores.



Colours indicated in the legend but not visible in the plots correspond to curves being perfectly congruent to those obtained for that variable from one or more of the other LOO splits.

Figure S7 Response curves for *Chaetocnema danielssoni*. For *Chaetocnema danielssoni*, marginal response curves obtained from the tuned Maxent and RF models fitted for each train-test leave-one-out (LOO) split, for the top three variables in terms of weighted average standardized importance scores.

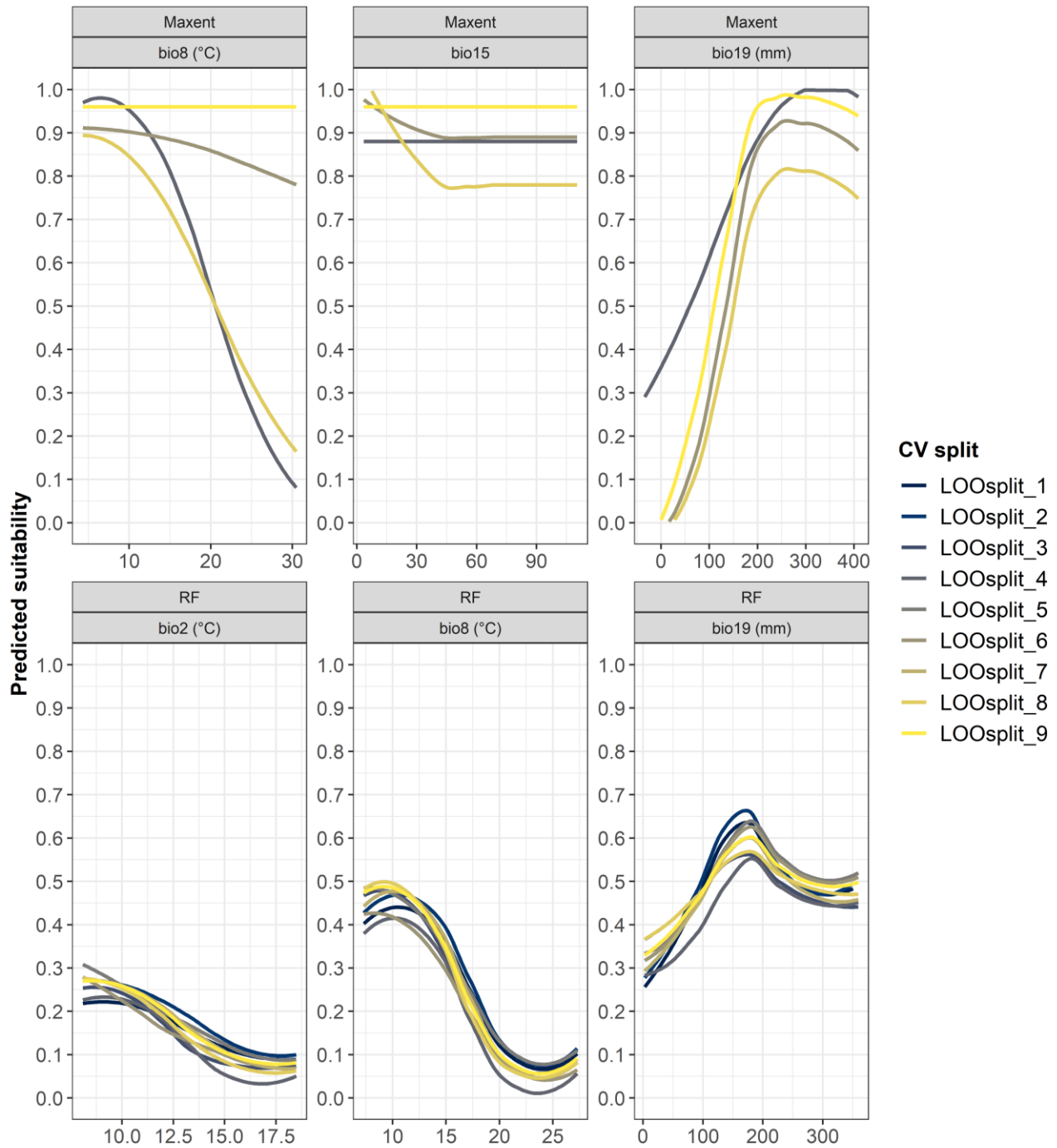


Figure S8 Response curves for *Chaetocnema darwini*. For *Chaetocnema darwini*, marginal response curves obtained from the tuned Maxent and RF models fitted for the two Checkerboard CV splits, for the top three variables in terms of weighted average standardized importance scores.

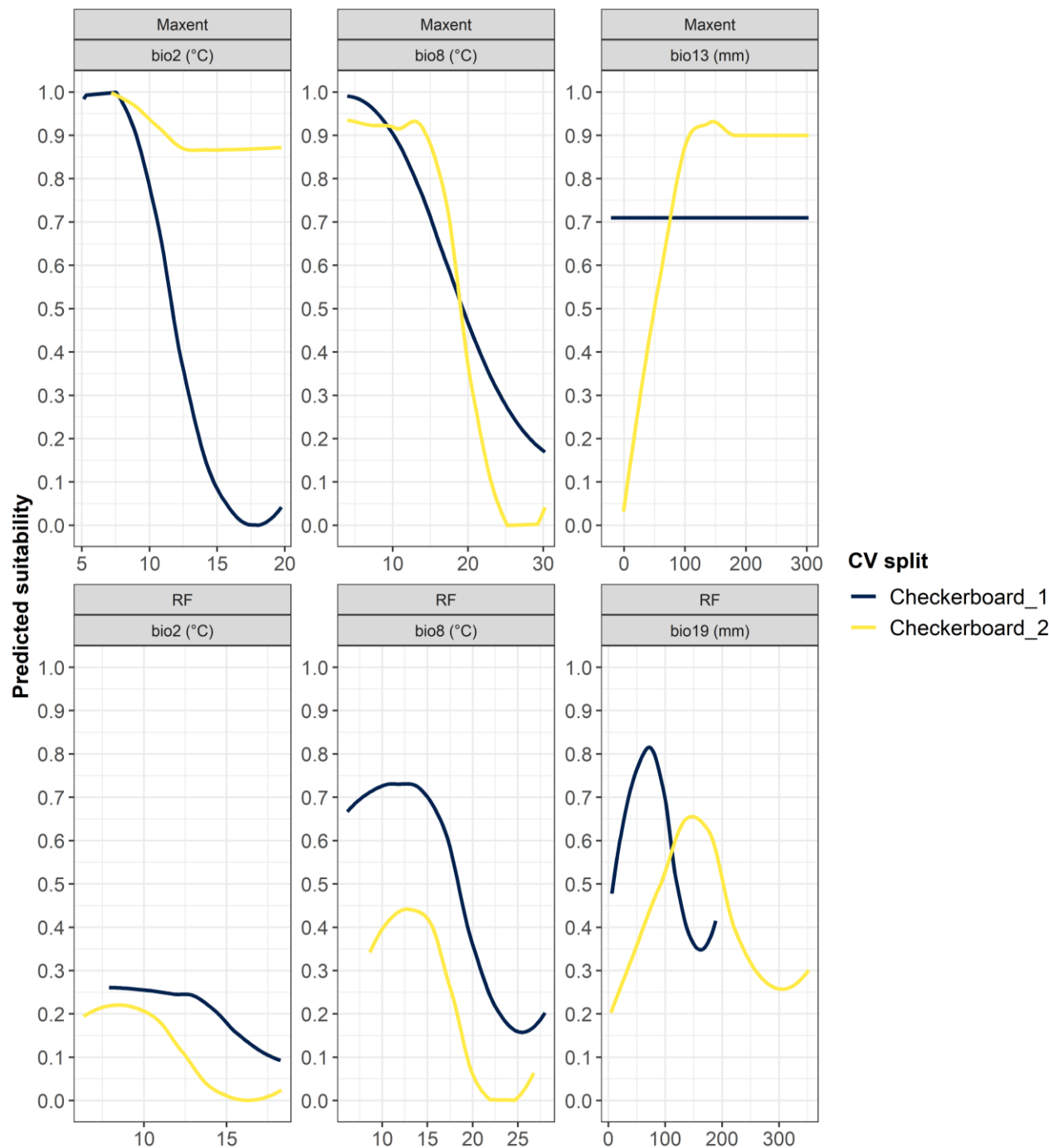
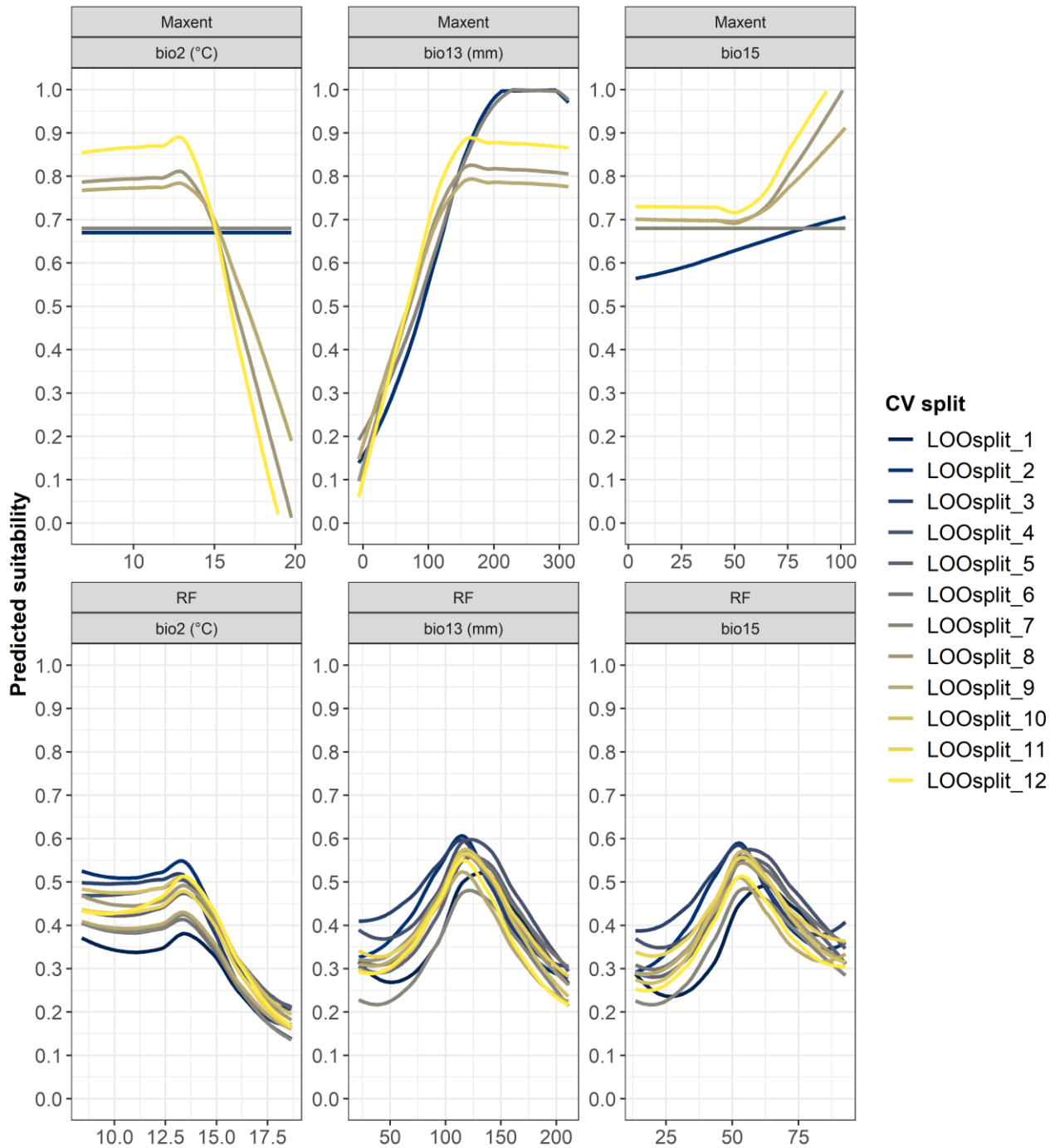


Figure S9 Response curves for *Chaetocnema gahani*. For *Chaetocnema gahani*, marginal response curves obtained from the tuned Maxent and RF models fitted for each train-test leave-one-out (LOO) split, for the top three variables in terms of weighted average standardized importance scores.



Colours indicated in the legend but not visible in the plots correspond to curves being perfectly congruent to those obtained for that variable from one or more of the other LOO splits.

Figure S10 Response curves for *Chaetocnema natalensis*. For *Chaetocnema natalensis*, marginal response curves obtained from the tuned Maxent and RF models fitted for the two Checkerboard CV splits, for the top three variables in terms of weighted average standardized importance scores.

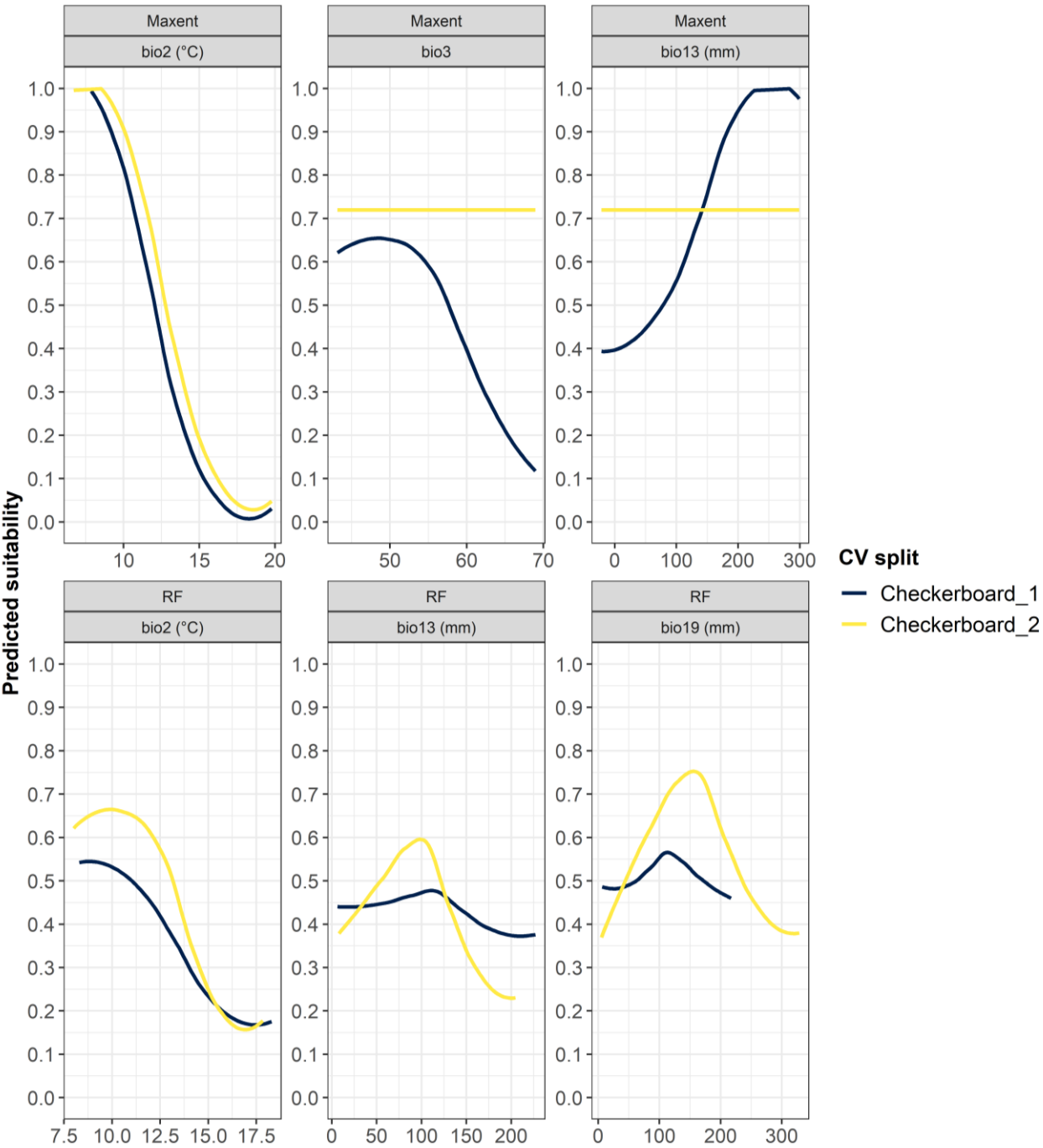
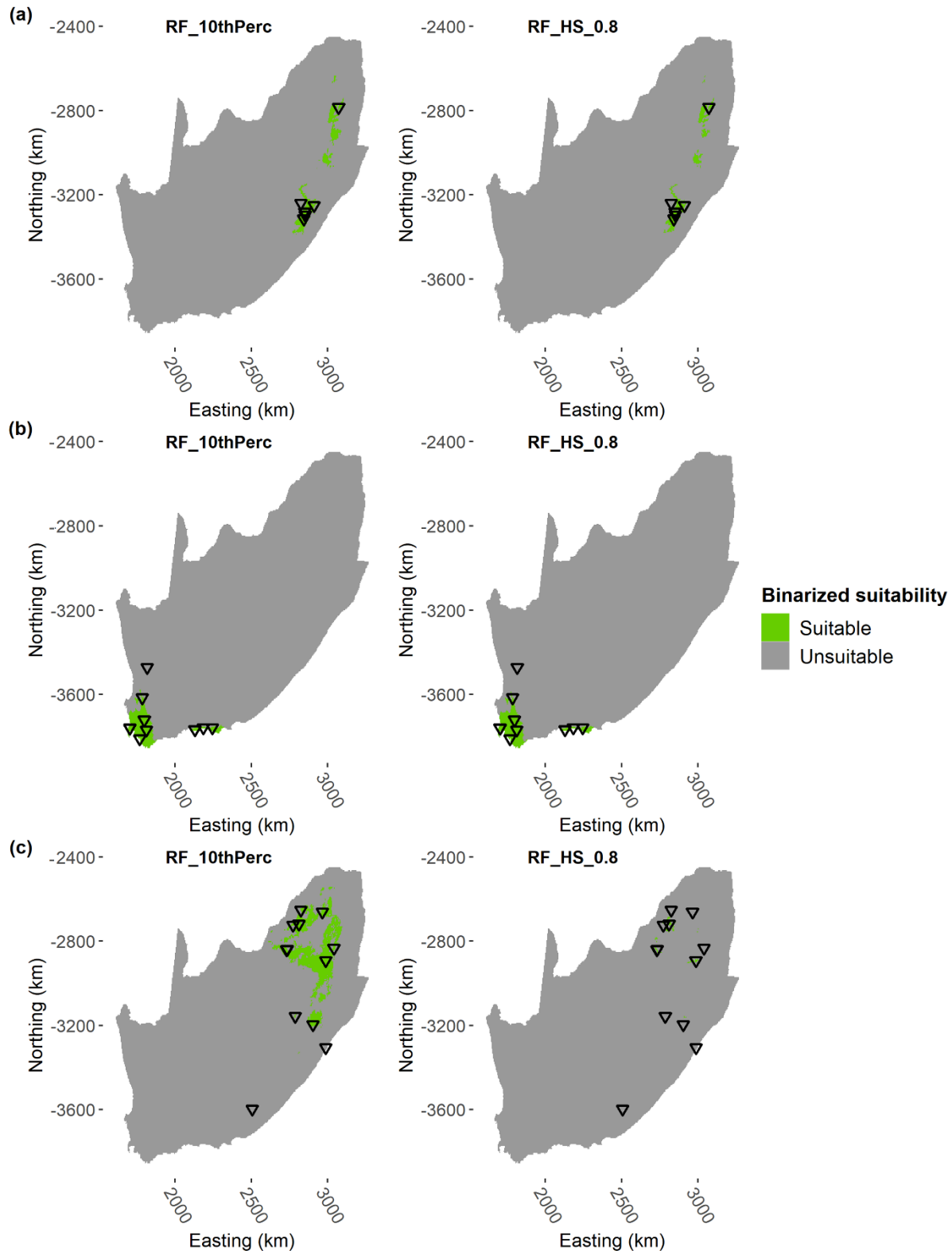


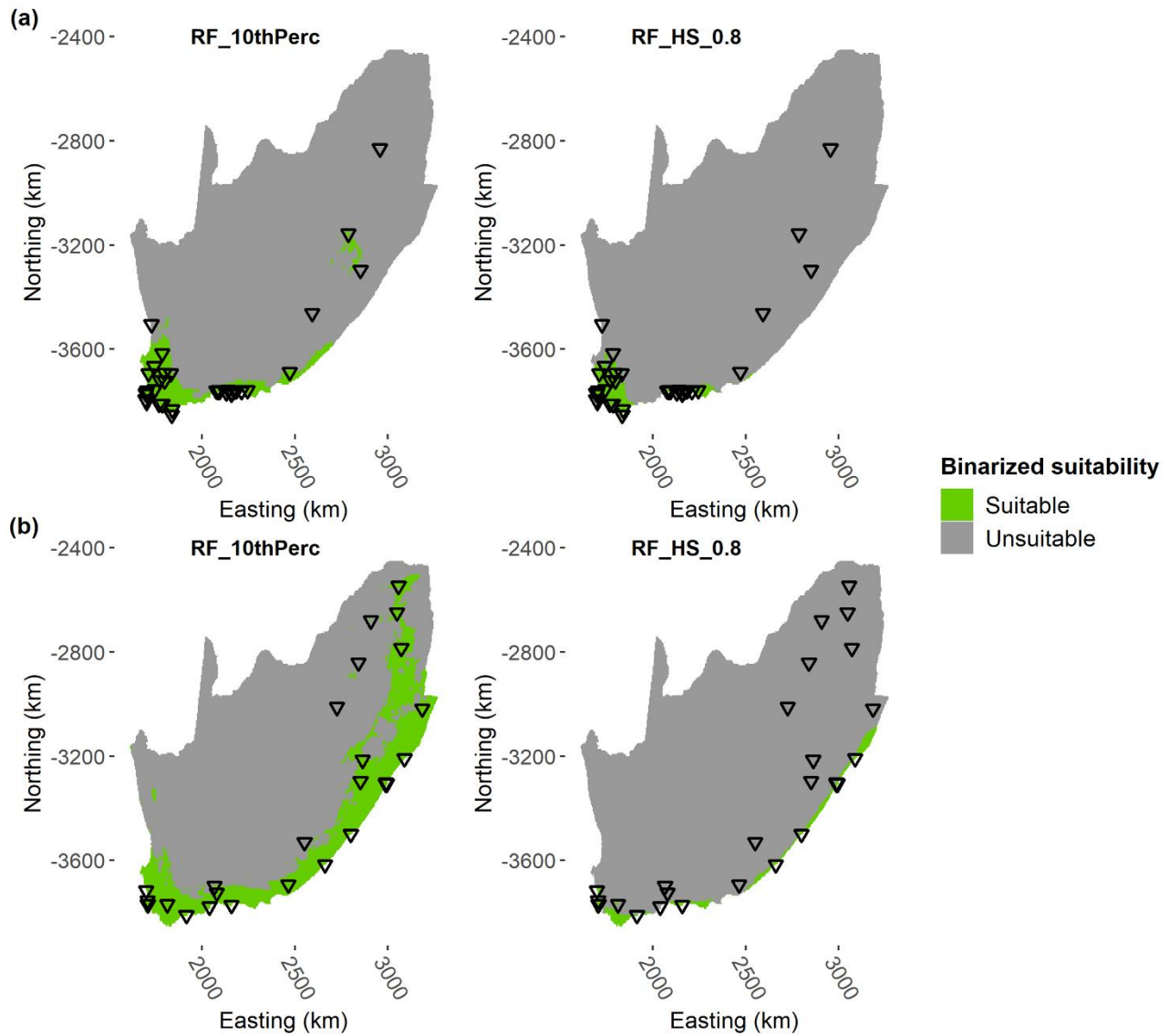
Figure S11 Binarized suitability for *Chaetocnema brincki*, *C. danielssoni* and *C. gahani*.

Binarized suitability resulting from the conversion of the RF-derived Wgt_Avg HS to binary predictions based on 10th percentile training presence threshold (left maps) or on a threshold corresponding to HS ≥ 0.8 (right maps), for: (a) *Chaetocnema brincki*; (b) *C. danielssoni*; (c) *C. gahani*.



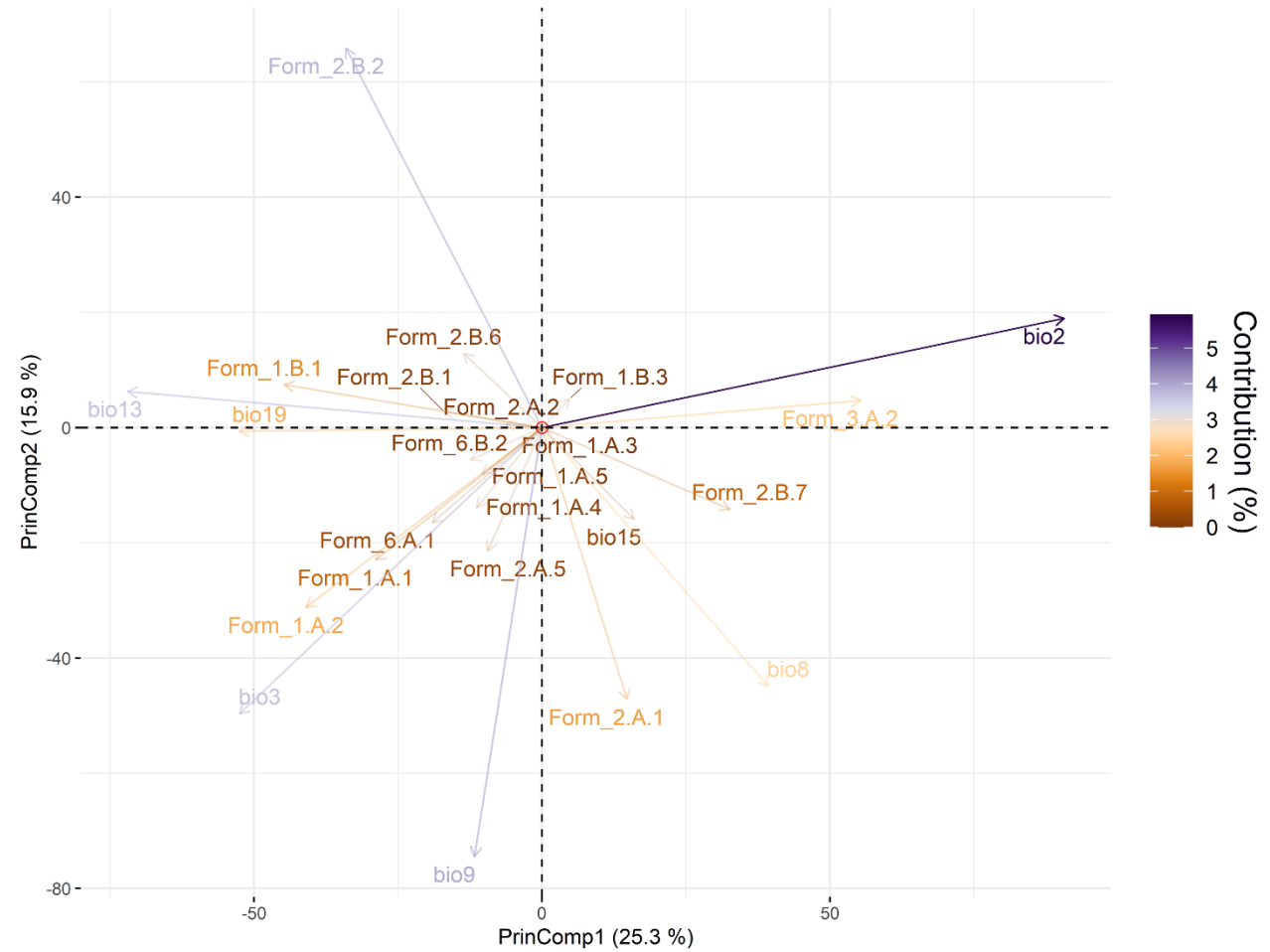
Reversed black-contoured triangles indicate the species' occurrence localities.

Figure S12 Binarized suitability for *Chaetocnema darwini* and *C. natalensis*. Binarized suitability resulting from the conversion of the RF-derived Wgt_Avg HS to binary predictions based on 10th percentile training presence threshold (left maps) or on a threshold corresponding to HS ≥ 0.8 (right maps), for: (a) *Chaetocnema darwini*; (b) *C. natalensis*.



Reversed black-contoured triangles indicate the species' occurrence localities.

Figure S13 Variables contribution to Principal Components form PCA-Env. Percent contribution of the bioclimatic variables selected for model fitting, and of the percent coverage from the 17 vegetation formations (see Table 1), to the first two Principal Components derived from a PCA-Env analysis summarizing environmental heterogeneity across the study region.



Relative contribution of each variable is shown in an orange-purple colour scale.