Geospatial modeling approach to monument construction using Michigan from A.D. 1000-1600 as a case study

Howey, M.C.L.; Palace, M.W.; McMichael, C.H.

DOI
10.1073/pnas.1603450113

Publication date
2016

Document Version
Other version

Published in
Proceedings of the National Academy of Sciences of the United States of America

License
Other

Supporting Information

Howey et al. 10.1073/pnas.1603450113

SI Text

Nominal Regression. We conducted a stepwise nominal regression to serve as a nonspatial, independent analysis of the power of the input variables used in MaxEnt modeling to discriminate between mounds and enclosures; this was done to help control for the tendency of MaxEnt to overestimate geospatial distributions. Analysis was done in JMP. The overall results show that the input variables do separate the two features (Fig. S1).

The regression also identified three variables as sufficient to separate the features statistically: distance to a river, SRTM topography, and BIO8, which is the mean temperature of the wettest quarter [which is the same 3-mo segment as BIO10 (mean temperature of warmest quarter]. These three variables align with the variables found to be key in predicting mound and enclosure differential landscape distribution in MaxEnt (Fig. S2).

Multivariate Correlation Table. We computed a multivariate correlation table for 13 input variables derived from 100,000 randomly generated points placed across the study area (Table S1). Table S1 provides a means of assessing linear relationships between each covariate used in the modeling. Of important note, distance to a lake and distance to a river have very low correlation coefficients across all variables. Analysis was done in JMP (Table S1).

Mound MaxEnt Model AUC Analysis. The test omission rate and predicted area as a function of the cumulative threshold averaged over the replicate runs were computed for mounds and are presented in Fig. S3.

The receiver operating characteristic curve was produced with the same data, again averaged over the replicate runs. Note that the specificity is defined using predicted area rather than true commission (40). The average test AUC for the replicate runs is 0.758, and the SD is 0.059; this is shown in Fig. S4.

Earthwork Enclosure MaxEnt Model AUC Analysis. The test omission rate and predicted area as a function of the cumulative threshold averaged over the replicate runs were computed for earthworks and are presented in Fig. S5.

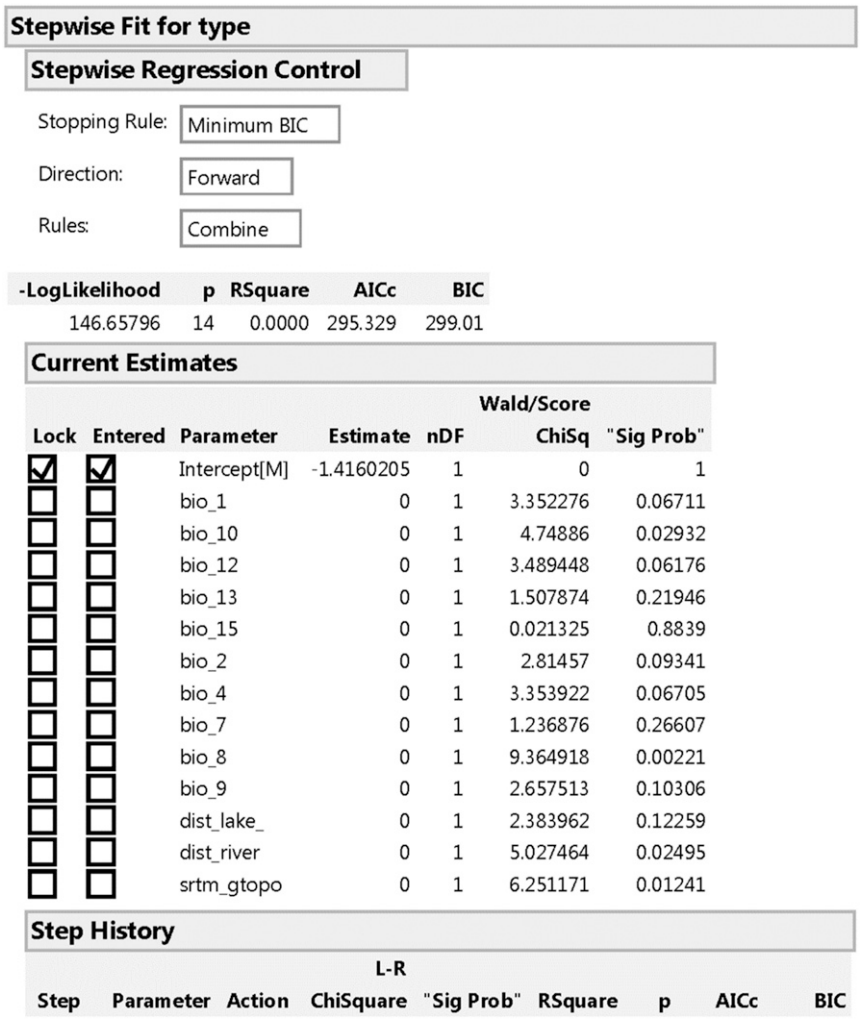
The receiver operating characteristic curve was produced with the same data, again averaged over the replicate runs. Note that the specificity is defined using predicted area rather than true commission (40). The average test AUC for the replicate runs is 0.852, and the SD is 0.031; this is shown in Fig. S6.

Mound MaxEnt Model Variable Contribution Jackknife Tests. Jackknife training and test gain assessments were done for mounds. Training gain values are derived from points used to train the model, and test gain values are derived from occurrence points used to test the predictive capacity of the model. Training and testing gain values with variable omitted are gain values of the model when the variable is excluded, and only variable is the gain value when the variable is used as a single predictor (Fig. S7).

Earthwork Enclosure MaxEnt Model Variable Contribution Jackknife Tests. Jackknife training and test gain assessments were done for earthwork enclosures. Training gain values are derived from points used to train the model, and test gain values are derived from occurrence points used to test the predictive capacity of the model. Training and testing gain values with variable omitted are gain values of the model when the variable is excluded, and only variable is the gain value when the variable is used as a single predictor (Fig. S8).

Mound Variable Response Curves. Response curves were created using each input variable alone for the MaxEnt model for mounds, which shows the dependence of predicted suitability as it patterns with the selected variable. The curves show the mean response of the replicate MaxEnt runs (red in Fig. S9) and ±1 SD (blue in Fig. S9). All temperature variables are shown in degrees Celsius by a factor of 10, all precipitation variables are in centimeters, distance to a lake and distance to a river are in kilometers and SRTM topography is in meters (Fig. S9).

Earthwork Enclosure Variable Response Curves. Response curves were created using each input variable alone for the MaxEnt model for enclosures, which shows the dependence of predicted suitability as it patterns with the selected variable. The curves show the mean response of the replicate MaxEnt runs (red in Fig. S10) and ±1 SD (blue in Fig. S10). All temperature variables are shown in degrees Celsius by a factor of 10, all precipitation variables are in centimeters, distance to a lake and distance to a river are in kilometers, and SRTM topography is in meters (Fig. S10).



Stepwise Fit for type

Stepwise Regression Control

Stopping Rule: Minimum BIC

Direction: Forward

Rules: Combine

-LogLikelihood	p	RSquare	AICc	BIC
146.65796	14	0.0000	295.329	299.01

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	Wald/Score ChiSq	"Sig Prob"
☑	☑	Intercept[M]	-1.4160205	1	0	1
☐	☐	bio_1	0	1	3.352276	0.06711
☐	☐	bio_10	0	1	4.74886	0.02932
☐	☐	bio_12	0	1	3.489448	0.06176
☐	☐	bio_13	0	1	1.507874	0.21946
☐	☐	bio_15	0	1	0.021325	0.8839
☐	☐	bio_2	0	1	2.81457	0.09341
☐	☐	bio_4	0	1	3.353922	0.06705
☐	☐	bio_7	0	1	1.236876	0.26607
☐	☐	bio_8	0	1	9.364918	0.00221
☐	☐	bio_9	0	1	2.657513	0.10306
☐	☐	dist_lake_	0	1	2.383962	0.12259
☐	☐	dist_river	0	1	5.027464	0.02495
☐	☐	srtm_gtopo	0	1	6.251171	0.01241

Step History

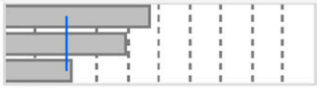
Step	Parameter	Action	L-R ChiSquare	"Sig Prob"	RSquare	p	AICc	BIC

Fig. S1. Overall nominal regression results. AICc, corrected Akaike information criterion; BIC, Bayesian information criterion; nDF, degrees freedom.

Nominal Logistic Fit for type

Effect Summary

Source	LogWorth	PValue
srtm_gtopo	4.701	0.00002
bio_8	3.908	0.00012
dist_river	2.186	0.00652



Converged in Gradient, 6 iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	15.43059	3	30.86119	<.0001*
Full	131.22737			
Reduced	146.65796			

RSquare (U)	0.1052
AICc	270.592
BIC	285.23
Observations (or Sum Wgts)	297

Measure	Training	Definition
Entropy RSquare	0.1052	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.1573	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.4418	$\sum$ -Log(ρ[j])/n
RMSE	0.3766	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.2837	$\sum \lvert y[j]-\rho[j]\rvert/n$
Misclassification Rate	0.1987	$\sum$ (ρ[j]≠ρMax)/n
N	297	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	284	127.06849	254.137
Saturated	287	4.15888	Prob>ChiSq
Fitted	3	131.22737	0.8982

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-8.9351516	1.8161754	24.20	<.0001*
bio_8	0.02514725	0.0069351	13.15	0.0003*
dist_river	-14.009887	5.7962083	5.84	0.0156*
srtm_gtopo	0.0127345	0.0031639	16.20	<.0001*

For log odds of E/M

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
bio_8	1	1	14.7383869	0.0001*
dist_river	1	1	7.40155449	0.0065*
srtm_gtopo	1	1	18.197676	<.0001*

Fig. S2. Variable nominal regression results. AICc, corrected Akaike information criterion; BIC, Bayesian information criterion; E/M, log odds from enclosures/mounds; RMSE, root mean squared error.

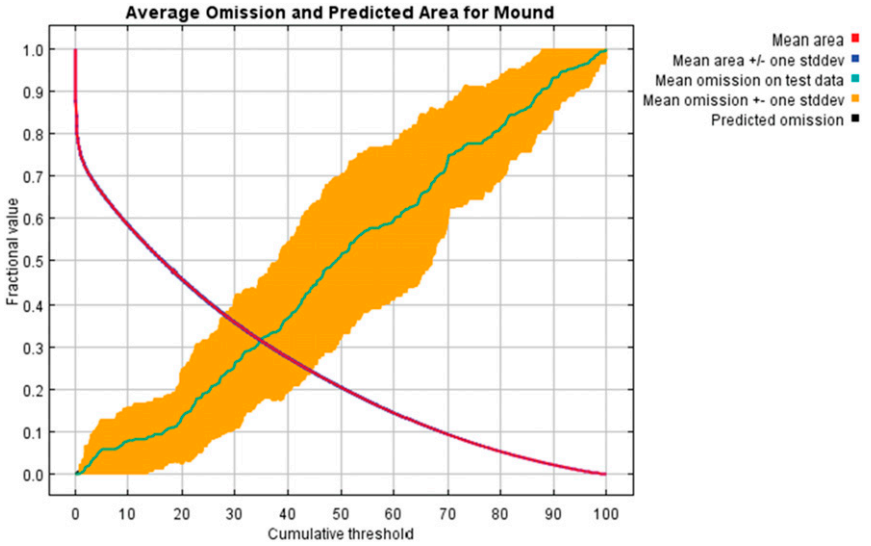


Fig. S3. Mound average omission and predicted area.

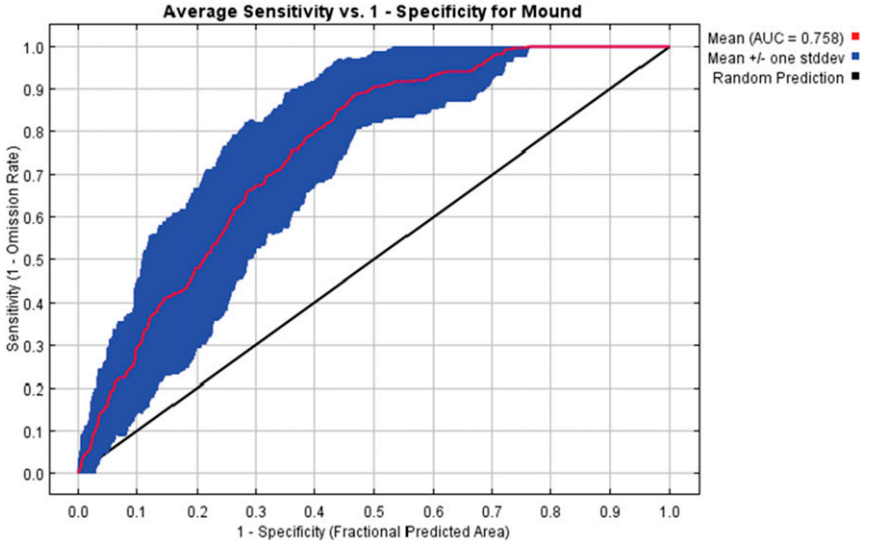


Fig. S4. Mound receiver operating characteristic curve.

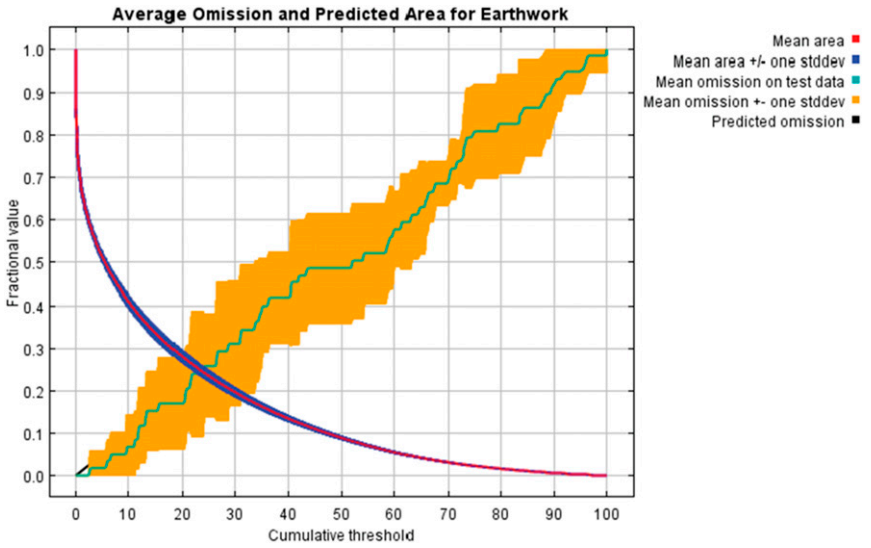


Fig. S5. Enclosure average omission and predicted area.

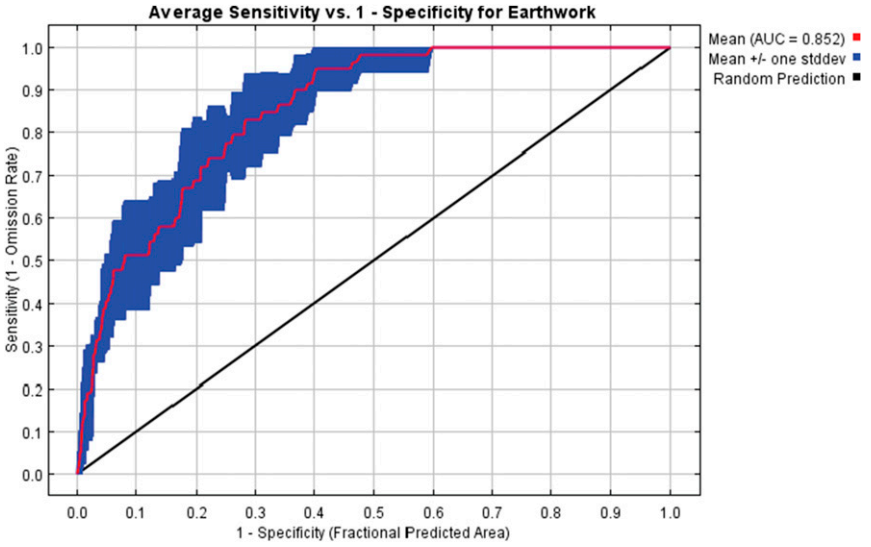


Fig. S6. Enclosure receiver operating characteristic curve.

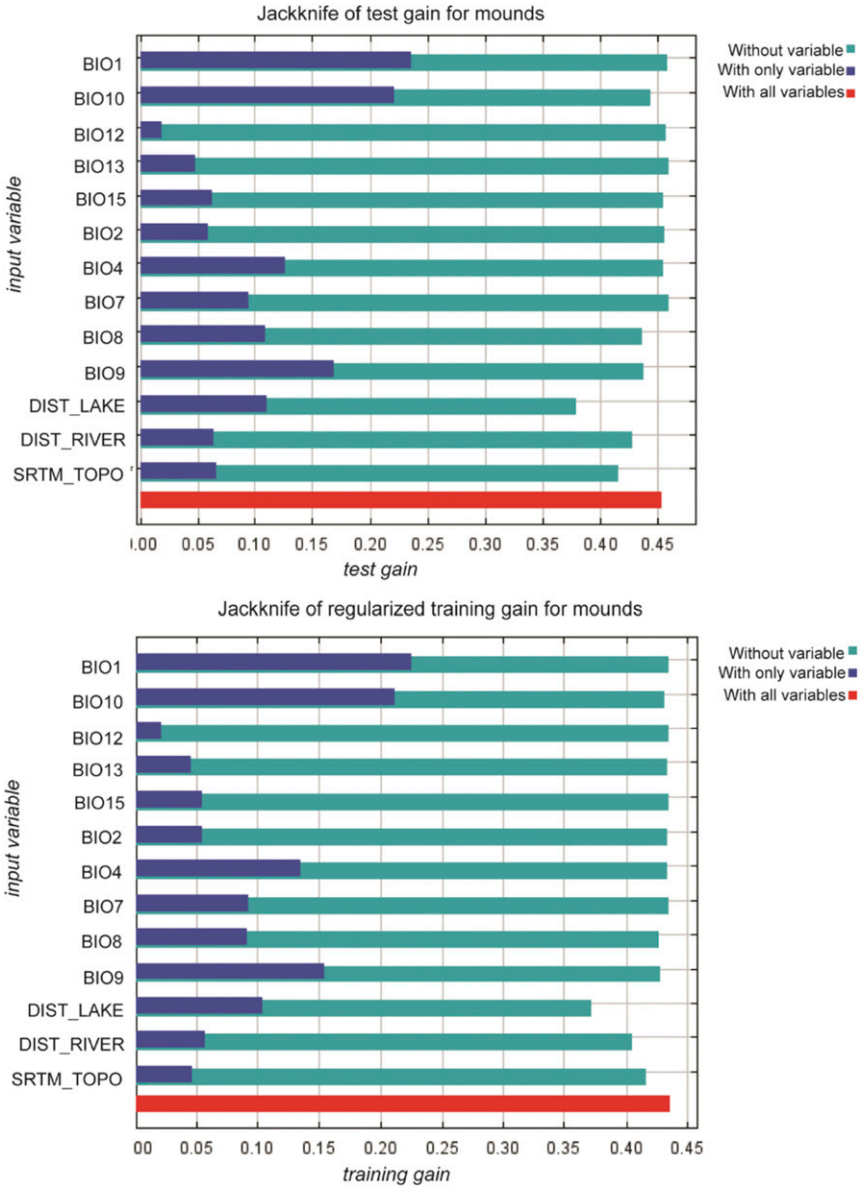


Fig. S7. Mound jackknife tests gain. Topo, topography.

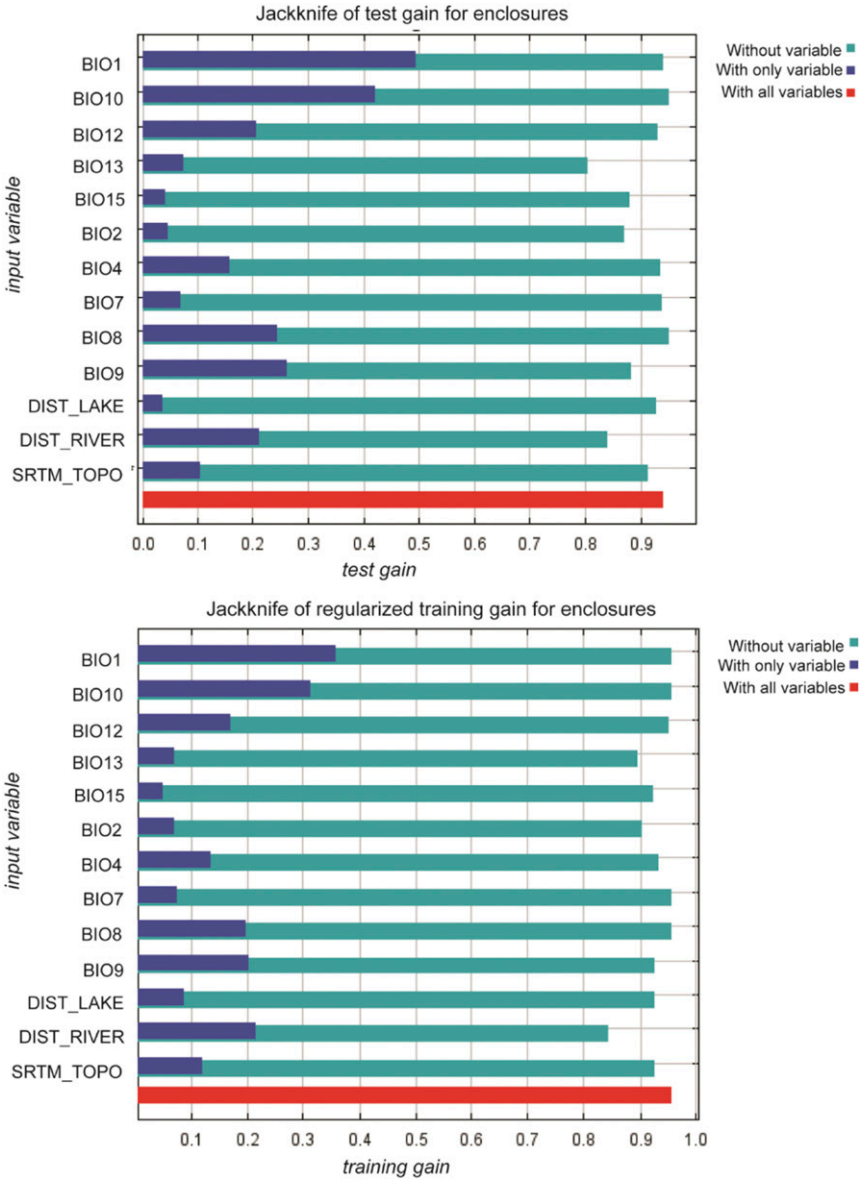


Fig. S8. Enclosure jackknife tests. Topo, topography.

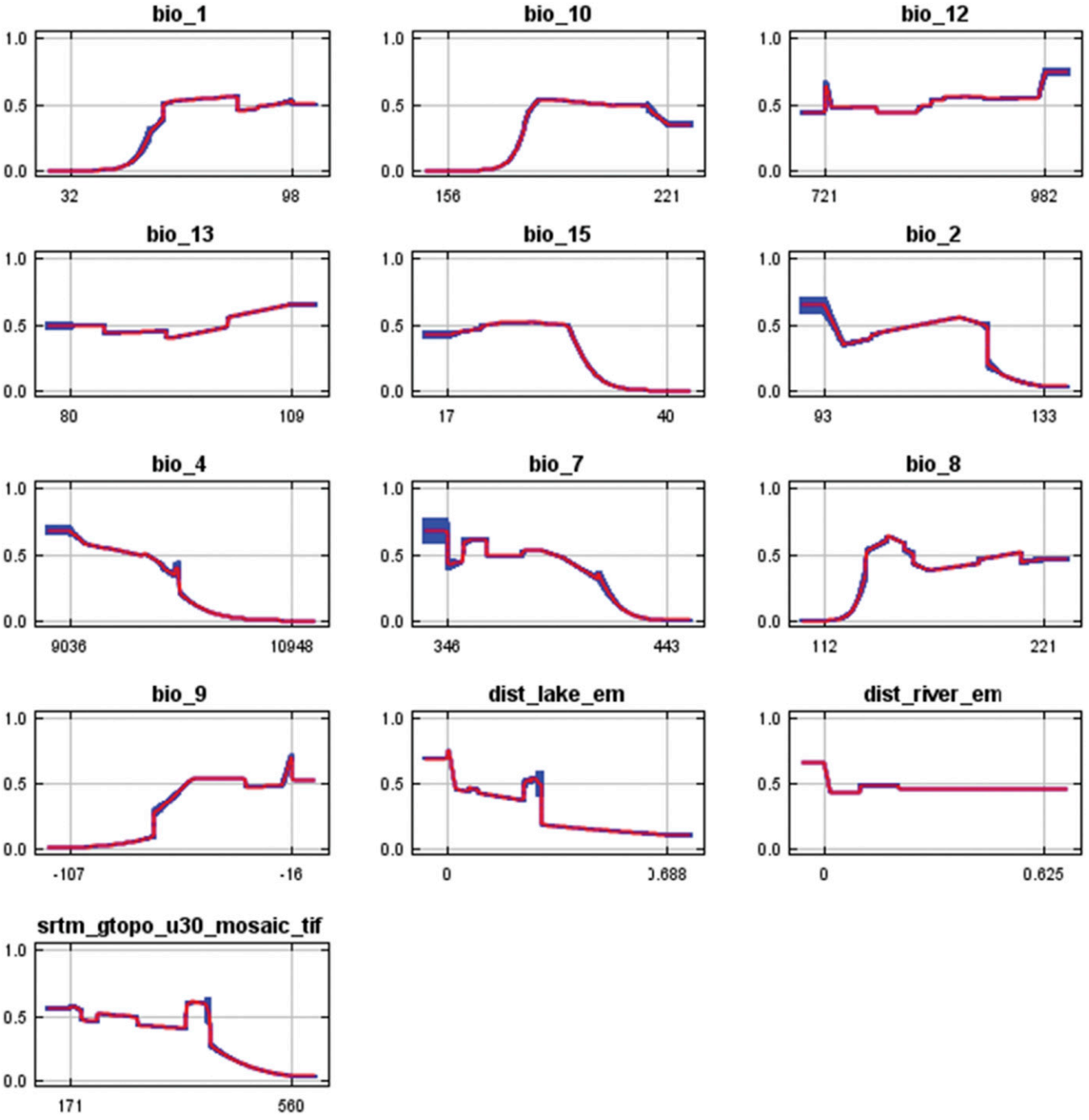


Fig. S9. Mound variable response curves.

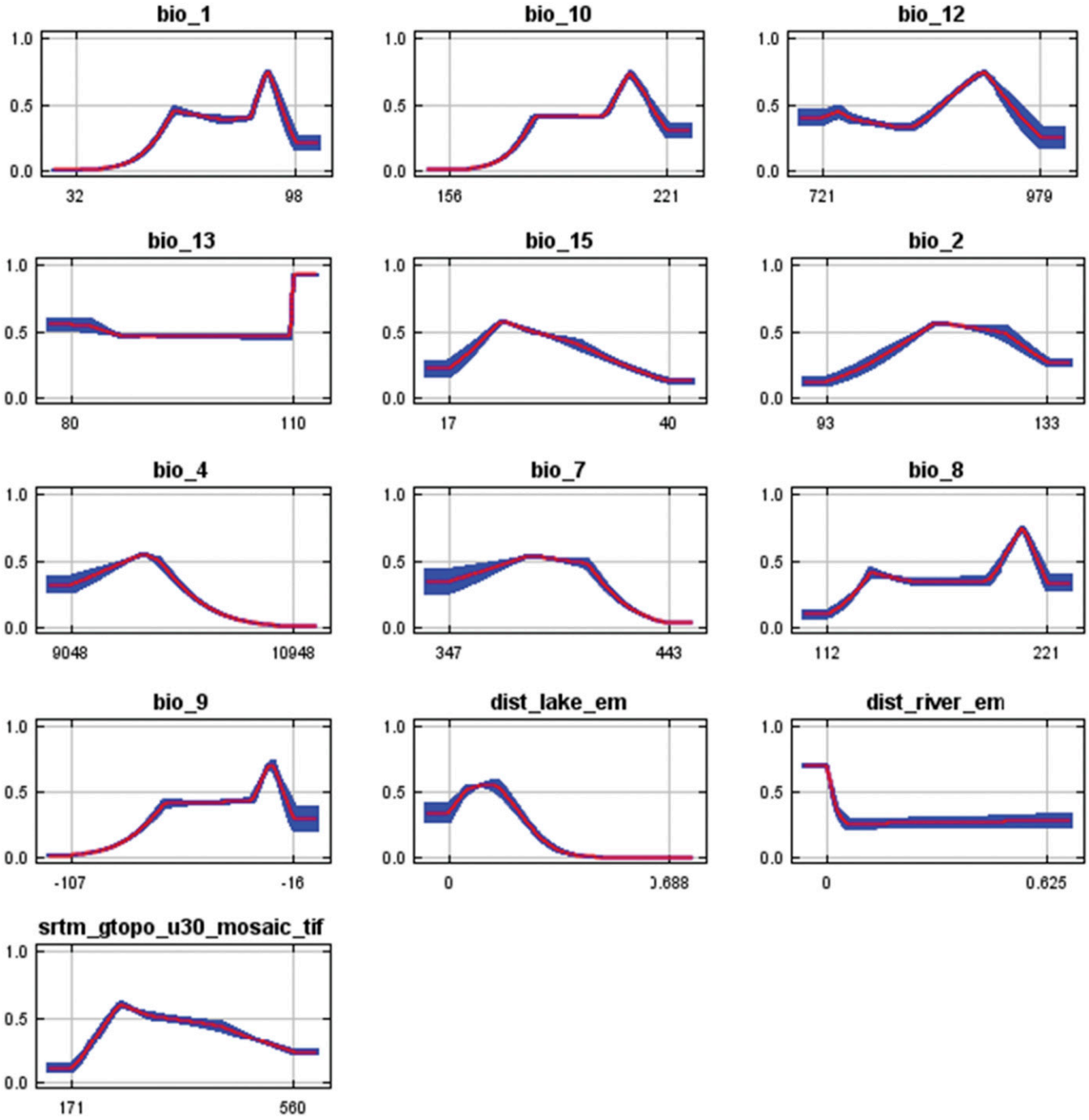


Fig. S10. Enclosure variable response curves.

Table S1. Multivariate correlation table

	BIO1	BIO10	BIO12	BIO13	BIO15	BIO2	BIO4	BIO7	BIO8	BIO9	Dist_lake	Dist_river	Topo
BIO1	1.000	0.979	0.258	−0.203	−0.561	−0.392	−0.721	−0.738	0.667	0.791	0.184	−0.125	−0.514
BIO10	0.979	1.000	0.238	−0.225	−0.454	−0.276	−0.569	−0.604	0.755	0.713	0.191	−0.155	−0.422
BIO12	0.258	0.238	1.000	0.549	−0.383	−0.110	−0.179	−0.254	0.014	0.333	−0.173	−0.006	0.078
BIO13	−0.203	−0.225	0.549	1.000	0.217	0.302	0.089	0.183	−0.488	−0.188	−0.278	0.061	0.335
BIO15	−0.561	−0.454	−0.383	0.217	1.000	0.749	0.733	0.825	−0.157	−0.792	−0.139	−0.114	0.608
BIO2	−0.392	−0.276	−0.110	0.302	0.749	1.000	0.657	0.831	−0.084	−0.601	−0.245	−0.171	0.701
BIO4	−0.721	−0.569	−0.179	0.089	0.733	0.657	1.000	0.942	−0.127	−0.773	−0.107	−0.024	0.660
BIO7	−0.738	−0.604	−0.254	0.183	0.825	0.831	0.942	1.000	−0.219	−0.823	−0.181	−0.041	0.697
BIO8	0.667	0.755	0.014	−0.488	−0.157	−0.084	−0.127	−0.219	1.000	0.362	0.246	−0.195	−0.167
BIO9	0.791	0.713	0.333	−0.188	−0.792	−0.601	−0.773	−0.823	0.362	1.000	0.125	−0.043	−0.624
Dist_lake	0.184	0.191	−0.173	−0.278	−0.139	−0.245	−0.107	−0.181	0.246	0.125	1.000	−0.052	−0.236
Dist_river	−0.125	−0.155	−0.006	0.061	−0.114	−0.171	−0.024	−0.041	−0.195	−0.043	−0.052	1.000	−0.083
Topo	−0.514	−0.422	0.078	0.335	0.608	0.701	0.660	0.697	−0.167	−0.624	−0.236	−0.083	1.000

Topo, topography.