

Ecography

ECOG-04365

Frishkoff, L. O., Mahler, D. L. and Fortin, M.-J. 2019. Integrating over uncertainty in spatial scale of response within multispecies occupancy models yields more accurate assessments of community composition. – Ecography doi: [10.1111/ecog.04365](https://doi.org/10.1111/ecog.04365)

Supplementary material

Appendix 1.

Table A1. Simulation conditions of communities for single scale ssMSOM.

Component	Parameter	Detection Regime				
		Perfect	High	Low	Low-Env	Low-Env Variable
Detection	$\mu_{\alpha 0}$	100	0	-1	-1	-1
	$\sigma_{\alpha 0}$	0	1	1	1	1
	$\mu_{\alpha 1}$	0	0	0	1	1
	$\sigma_{\alpha 1}$	0	0	0	0	1
Occupancy	$\mu_{\beta 0}$	-1	-1	-1	-1	-1
	$\sigma_{\beta 0}$	1	1	1	1	1
	$\mu_{\beta 1}$	0.5	0.5	0.5	0.5	0.5
	$\sigma_{\beta 1}$	1	1	1	1	1
	$\sigma_{\gamma 0}$	0.1	0.1	0.1	0.1	0.1
	s	variable	variable	variable	variable	variable
	s	variable	variable	variable	variable	variable
Sample Size	N_{sp}	16	16	16	16	16
	N_{site}	50	50	50	50	50
	N_{visit}	3	3	3	3	3

Table A2. Simulation conditions of communities for species-specific ssMSOM.

Component	Parameter	Detection Regime				
		Perfect	High	Low	Low-Env	Low-Env Variable
Detection	$\mu_{\alpha 0}$	100	0	-1	-1	-1
	$\sigma_{\alpha 0}$	0	1	1	1	1
	$\mu_{\alpha 1}$	0	0	0	1	1
	$\sigma_{\alpha 1}$	0	0	0	0	1
Occupancy	$\mu_{\beta 0}$	-1	-1	-1	-1	-1
	$\sigma_{\beta 0}$	1	1	1	1	1
	$\mu_{\beta 1}$	0.5	0.5	0.5	0.5	0.5
	$\sigma_{\beta 1}$	1	1	1	1	1
	$\sigma_{\gamma 0}$	0.1	0.1	0.1	0.1	0.1
	μ_{scale}	10	10	10	10	10
	σ_{scale}	5	5	5	5	5
	η	3	3	3	3	3
Sample Size	N_{sp}	16, 64	16, 64	16, 64	16, 64	16, 64
	N_{site}	120	120	120	120	120
	N_{visit}	3	3	3	3	3

Supplemental Figures:

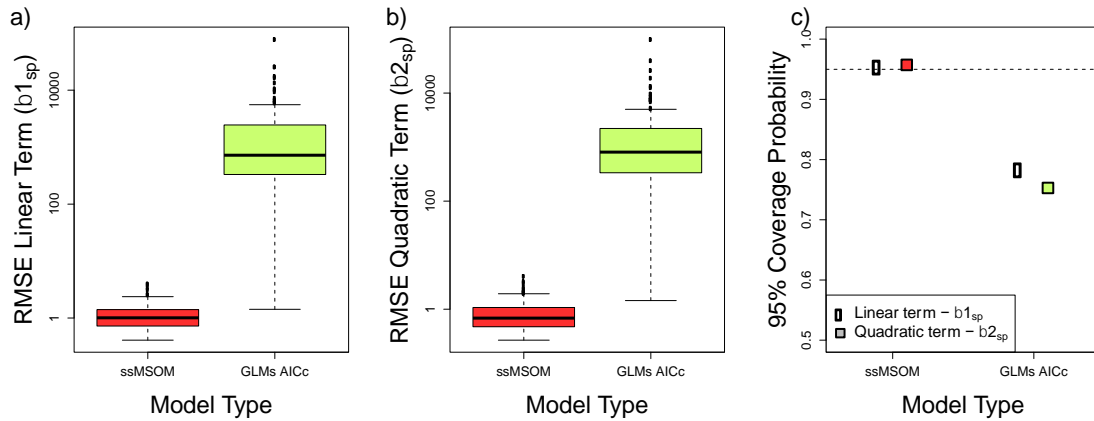


Figure A1. In empirical systems, species may respond to environmental gradients in non-linear fashions. We test whether the ssMSOM performs well for non-linear problems by examining a situation in which species are simulated, and modeled, according to a quadratic function of the environmental covariate. When estimating species' responses to the environment, the ssMSOM shows lower error for both linear (a) and quadratic (b) terms than GLMs. (c) The probability that the simulated value is within the 95% confidence intervals is at its expected value for the ssMSOM, but depressed for GLMs. Results presented here come from the "low" detection scenario (average detection probability of 0.25).

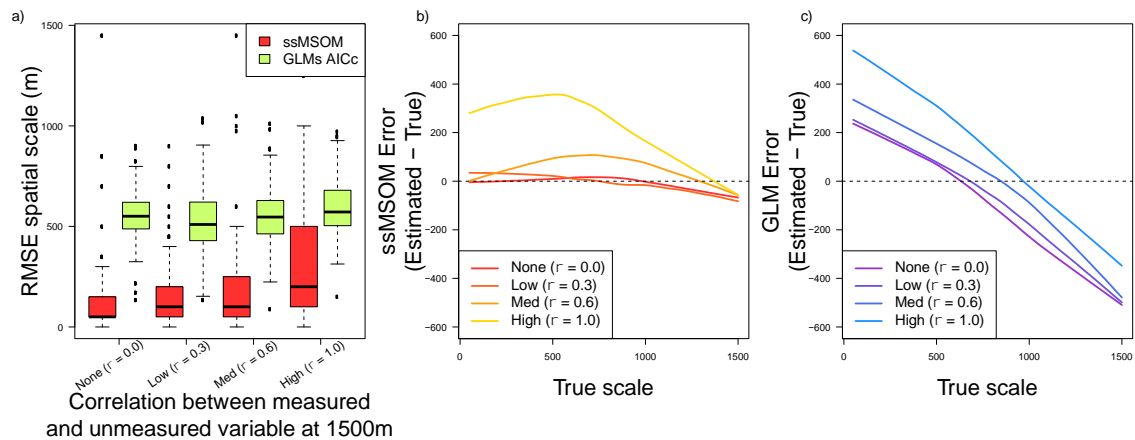


Figure A2. In empirical studies, environmental variables that affect community composition may be unknown or unmeasured. If the measured and unmeasured environmental variables are correlated, and if the scale at which unmeasured covariates affect the community is different from that of the measured covariates, then bias may be introduced into the scale estimates. For a single detection scenario (low), we simulated 120 datasets, varying the scale at which a measured environmental variable affected the community, but also including a second unmeasured environmental variable that affects communities at a scale of 1500m. In this simulation the measured and unmeasured covariates affect species with the same average strength (both β_1 and the effect of the unmeasured variable were drawn from a normal distribution with a mean of 0.5 and a SD of 1). After simulation an ssMSOM (or a series of GLMs) was run, without this second environmental covariate included in the model (corresponding to a scenario in which unmeasured variables affect community composition). In (a), root mean square error of the estimated spatial scale increases with the degree of correlation between environmental covariates. In (b) and (c), lines represent LOWESS smoothers for the degree of error depending on the spatial scale at which the measured environmental covariate affects the community for either the ssMSOM (b), or GLMs (c). Appreciable bias in scale estimation for the ssMSOM was only introduced when the second covariate was strongly correlated with the first, and when the scale of response for the first variable was different from the second (i.e. not close to 1500m). GLMs are prone to high error, though inclusion of an unmeasured variable consistently shifts estimates of spatial scale for the measured variable towards that of the unmeasured variable.

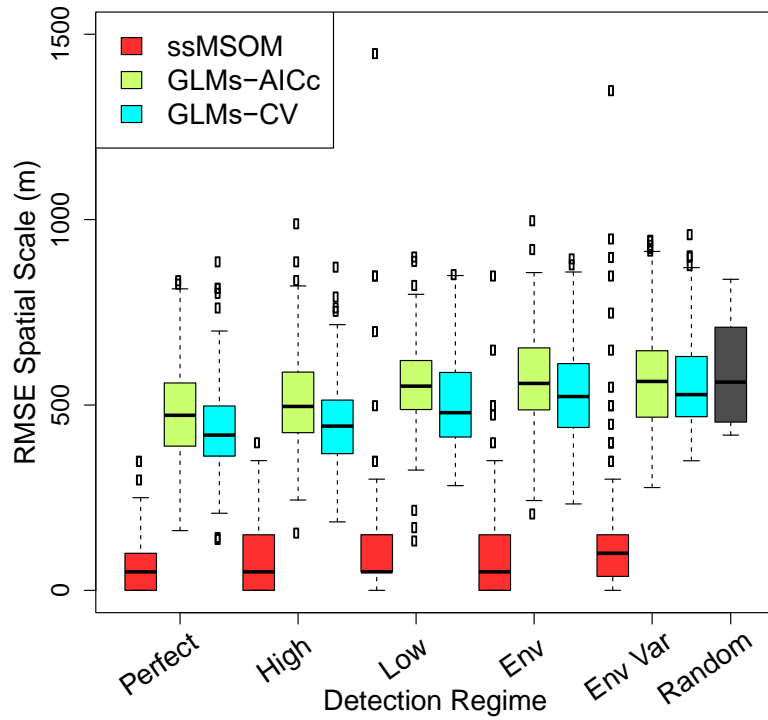


Figure A3. Comparison of performance of scale selection through ssMSOM with a scale selection parameter, versus scale selecting through choosing the GLM with the lowest AICc, versus choosing the GLM based on the lowest mean square prediction error from 5-fold cross validation (CV). GLMs with $p < 0.05$ correspond to a situation in which all species with non-significant β_1 terms were dropped from the analysis. Cross validation and out-of-sample predictive power do not lead to different conclusion from those of in-sample AICc based methods.

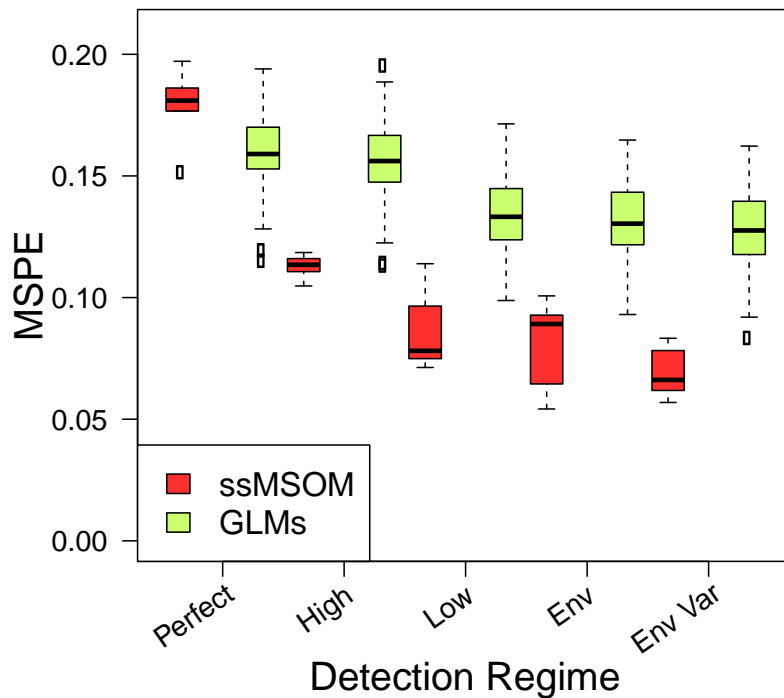


Figure A4. Mean square prediction error from 5-fold cross validation of ssMSOM versus GLMs. For ssMSOM, MSPE was calculated for a single model run, while for GLMs the presented MSPE is from the scale for which MSPE is minimized. Because of long cross-validation run times, only 5 simulated datasets were cross-validated for the ssMSOM for each detection scenario. Cross validation in contrast is rapid for GLMs, and so all species for all 240 simulated datasets are presented.

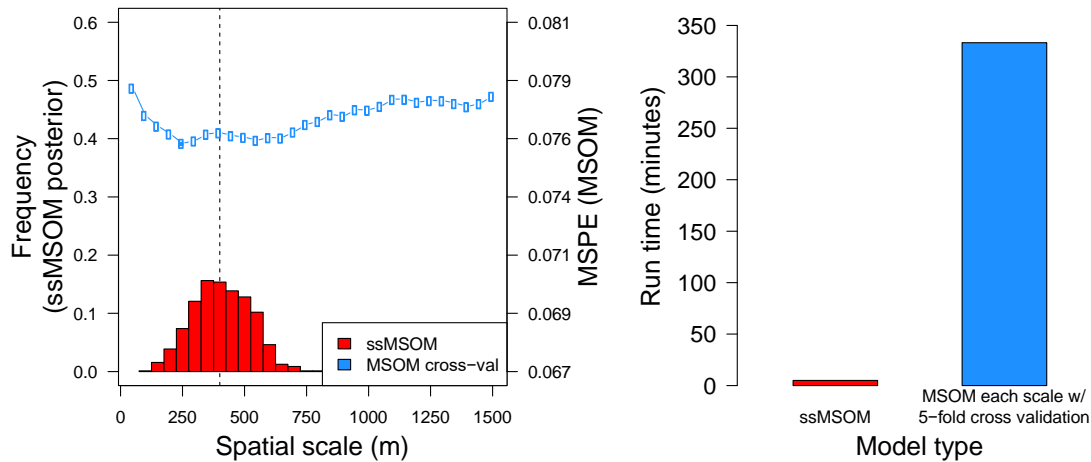


Figure A5. Comparison of ssMSOM versus MSOM run on each scale, with 5-fold cross validation to select scale that minimizes mean square predictive error (MSPE). In the left panel, the histogram shows the full posterior of scales supported by the data from ssMSOM, while points show MSPE for a MSOM run at that scale, with the scale that minimize MSPE represented by a filled circle. A dashed line depicts the true scale under which the data were simulated. In the right panel the total computer run time (on a 2.8 Ghz processor) to complete the analysis depicted in the left panel. Both the ssMSOM and the MSOMs with cross-validation support roughly the same range of scales, however the ssMSOM does so in 5 minutes, as opposed to 5.5 hours.