Appendix B. Cade, B.S. 2015. Model averaging and muddled multimodel inferences. *Ecology*

Simulations of multi-part compositional predictors within a zero-truncated Poisson regression model with proportions of cover types and correlations similar to those in the breeding sage-grouse location count model of Rice et al. (2013) and an explanation and evaluation of how they used predicted mean counts to predict probability of occupancy.

Simulations of model-averaged compositional predictors

Because the data used by Rice et al. (2013) were not available, I demonstrate the properties of model-averaged estimates for compositional predictors by simulating data that is similar to the pattern Rice et al. (2013) observed for breeding sage-grouse. Here I combine the original 13-part composition of cover type proportions in Rice et al. (2013) into a simpler 5-part composition that provides a more tractable number of candidate models to estimate. I combined all the shrub cover types (sagebrush, mountain shrub, salt desert shrub, and shrubland) into one cover type designated sagebrush/shrub (X_1) which is structured to vary uniformly from 0.50-0.95. All herbaceous cover types likely to be used by sage-grouse (grassland, riparian, and agriculture) in the breeding season were combined into one cover type designated herbaceous (X_2) that varies uniformly from 0.00-0.30, with the constraint that the sum of the proportions of these two cover types ($X_1 + X_2$) range from 0.80-1.00, and that both these cover types were negatively correlated with the forest (X_3) and alpine (X_4) cover types (Table B1). The fifth part, proportion urban, was given as a proportion 0.00-0.02 and was excluded from all models similar to Rice et al. (2013). This yields a linear function of $X_1 + X_2$ that is linearly related to $X_3 + X_4$ with $r = -0.998$ for the sample of $n = 200$ 1-km^2 units used in estimation.

The R script in Supplement 2 was used to generate the five predictors with the unit sum constraint and to provide random variables from a Poisson regression model with $E[y|\mathbf{X}] = \exp(-5.8 + 6.3X_1 + 15.2X_2)$. The nonzero subset of random variables (n = 165) was retained for model estimation. I estimated all eight candidate models that included the sagebrush/shrub (X_1) predictor with a zero-truncated Poisson regression following the protocol of Rice et al. (2013), computed AIC weights, and obtained model-averaged estimates of parameters for all four predictors, where estimates of zero were used for models where the predictor was not included for estimation (Burnham and Anderson 2002, Lukacs et al. 2010). I did not use a mixed-effect model as Rice et al. (2013) did because the incorporation of random effects has no bearing on the issues with compositional predictor variables. I also computed variance inflation factors and partial standard deviations that account for the multicollinearity among predictors within each model and provided standardized estimates (based on transforming the X_i) and computed model-averaged standardized estimates.

My model-averaged estimates are positive for the predictors positively related to sage-grouse location counts and negative for those inversely related by definition of the composition (Table B2) similar to the pattern observed by Rice et al. (2013, Table 3). But note that some estimates for the sagebrush/shrub predictor (X_1) are negative even for models with substantial AIC weight (Table B2). The model-averaged estimates for both sagebrush/shrub (X_1) and herbaceous (X_2) predictors are severely attenuated compared to the parameter values because of the reductions in estimate size when the redundant predictors forest (X_3) and alpine (X_4) are included in models, with more pronounced deviations for the sagebrush/shrub predictor. The extreme differences among models in partial standard deviations for a given predictor indicate the major scale changes in units, precluding any meaningful use of simple model-averaged estimates. But in this example, even the standardized parameter estimates have wide variation, especially for the sagebrush/shrub predictor (Table B2) due to the effect of including redundant compositional predictors. Clearly, with this level of redundancy associated with compositional predictors, there must be some additional constraints imposed on the candidate model set considered, e.g., just one subset of predictors, either the sagebrush/shrub and herbaceous predictors or the forest and alpine predictors but not both. Both subsets of predictors might be incorporated simultaneously into model estimates if they are transformed with orthonormalizing log ratios (Aitchison and Egozcue 2005, Hron et al. 2012). Unfortunately, this does change the interpretation of the parameter estimates.

Predicting model-averaged mean counts from model-averaged regression coefficients

In this example with the given AIC weights among models, there is minimal difference among means estimated by model-averaging the predicted mean counts across the eight models (correct) compared to estimating model-averaged mean counts based on using model-averaged estimates for predictors (incorrect) across the observed range of predictor values (differences in predicted means range 0.00-0.025). However, this similarity or difference is a function of the difference in magnitude of estimates and their interactions with the AIC weights. For example, if I use the same model estimates as in Table B2 but assign them equal AIC weights (w_j = 1/8), then the difference between the correct and incorrect predictions of means are now substantially larger (differences in predicted means range 0.004-1.404).

Rice et al. (2013) also apparently used their model-averaged predictions of mean counts of telemetry locations to make predictions for the probability of any count ≥1, i.e., probability of occupancy. It is not clear why Rice et al. (2013) made this transformation nor do they discuss the consequences of shifting to predicting probabilities of occupancy. They apparently accomplished this by using the relationship between predicted means from a zero-truncated Poisson distribution and predicted means from a conventional Poisson distribution, which allowed for computing probabilities of counts ≥1 (W. Thogmartin, personal communication).

Given a zero-truncated Poisson distribution with mean $u_{ij}/(1 - \exp(-u_{ij}))$, the corresponding Poisson distribution with mean u_{ij} (using Rice et al. 2013 notation for individual i and population j) has probabilities of $y \geq 1$ equal to $P(y_{ij} \geq 1 | u_{ij}) = 1 - \exp(-u_{ij})$, a logistic response curve that transforms the exponential rates of change in mean counts to very different rates of change in the probability of occupancy scale (Figure B1). The response curve for Poisson probabilities of counts ≥1 has the inverse pattern of the response curve for means of the Poisson distribution associated with those probabilities, changing very rapidly for probabilities associated with means ≤2 and much less rapidly for probabilities associated with means >2. The implications of this are that the predicted probabilities of occupancy that were mapped in Rice et al. (2013, Figure 2) imply greater discrimination among areas without any sage-grouse than are reasonable to infer from a model estimated from samples only where sage-grouse occurred. Conversely, by shifting to predicting probabilities of occupancy, Rice et al. (2013) eliminated most of the response sensitivity to varying amounts of sage-grouse use that they actually modeled. Furthermore, their model validation was only performed on independent data where sage-grouse occurred (another telemetry data set and lek count data), providing no substantiation of the modeled relationships for areas unlikely to be occupied sage-grouse habitat which composed >50% of the mapped area for which predictions were made (Rice et al. 2013, compare Figures 1 and 2). Other aspects of their model validation exercise seemed questionable even for occupied sage-grouse areas, e.g., a mismatch between statistical units for predictions compared to observations for their rank correlations.

Other issues masked by model averaging

One of the unfortunate consequences of model averaging for multimodel inference is that people often fail to pay attention to important statistical details relevant to every single candidate model as noted by Giudice et al. (2012). There are several notable issues in the Rice et al. (2013) sage-grouse models that fall into that category, including poor use of random effects on just the intercept to account for repeated measures (Gillies et al. 2006, Schielzeth and Forstmeier 2008), basing predictions on a subject-specific effect rather than a population effect averaged across individual grouse (Fieberg et al. 2009), and failure to report sample size for any models. Addressing any or all of these issues would have been more productive for Rice et al. (2013) than trying to address multimodel inferences by model averaging.

LITERATURE CITED

Aitchison, J., and J. J. Egozcue. 2005. Compositional data analysis: Where are we and where should we be heading? Mathematical Geology 37:829–850.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Fieberg, J., R. H. Rieger, M. C. Zicus, and J. S. Schildcrout. 2009. Regression modelling of correlated data in ecology: subject specific and population averaged response patterns. Journal of Applied Ecology 46:1018–1025.

Gillies, C. S., M. Hebblewhite, S. E. Nielsen, M. A. Krawchuk, C. L. Aldridge, J. L. Frair, D. J. Saher, C. E Stevens, and C. L. Jerde. 2006. Application of random effects to the study of resource selection by animals. Journal of Animal Ecology 75:887–898.

Giudice, J. H., J. R. Fieberg, and M. S. Lenarz. 2012. Spending degrees of freedom in a poor economy: A case study of building a sightability model for moose in northeastern Minnesota. Journal of Wildlife Management 76:75–87.

Hron, K., P. Filzmoser, and K. Thompson. 2012. Linear regression with compositional explanatory variables. Journal of Applied Statistics 39:1115–1126.

Lukacs, P. M., K. P. Burnham, and D. R. Anderson. 2010. Model selection bias and Freedman's paradox. Annals of the Institute of Statistical Mathematics 62:117–125.

Rice, M. B., A. D. Appa, M. L. Phillips, J. H. Gammonley, B. P. Petch, and K. Eichoff. 2013. Analysis of regional species distribution models based on radio-telemetry datasets from multiple small-scale studies. Journal of Wildlife Management 77:821–831.

Schielzeth, H., and W. Forstmeier. 2008. Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology 145:416–420.

TABLE B1. Correlation matrix of simulated sage-grouse location counts (y) in 1-km^2 units from a Poisson distribution with a multi-part composition of proportion of cover types as predictors (n = 200); sagebrush/shrub (X_1), herbaceous (X_2), forest (X_3), and alpine (X_4); where the mean count was given by $E[y|\mathbf{X}] = \exp(-5.8 + 6.3X_1 + 15.2X_2)$. Simulation conditions were selected to be similar to those estimated for breeding sage-grouse in Rice et al. (2013).

	y	X_1	X_2	X_3	X_4
y	1.000	−0.081	0.705	−0.543	−0.507
X_1		1.000	−0.525	−0.498	−0.597
X_2			1.000	−0.419	−0.225
X_3				1.000	0.644
X_4					1.000

TABLE B2. Parameter estimates for the four predictors in eight models that included X_1, proportion sagebrush/shrub cover type, in the simulated zero-truncated Poisson model for sage-grouse location counts (n = 165). Unstandardized estimates and their standard errors, AIC weights w_j, and model-averaged estimates ignore the multicollinear covariance structure. Also provided are the inverse of the variance inflation factors VIF_{ij}^{-1}, partial standard deviations of X_{ij}, parameter estimates $\hat{\beta}_{ij}^{*}$ standardized by their partial standard deviations and their standard errors, and the model-averaged standardized parameter estimate and its standard error that account for scale changes associated with multicollinear predictors.

β_1 rate of change with proportion sagebrush/shrub (6.3)

Predictors used	Unstandardized estimates				Partial	Standardized estimates	
	$\hat{\beta}_{1j}$	$\widehat{\mathrm{se}}\,(\hat{\beta}_{1j} \mid g_j)$	w_j	VIF_{1j}^{-1}	SD X_{1j}	$\hat{\beta}_{1j}^{*}$	$\widehat{\mathrm{se}}\,(\hat{\beta}_{1j}^{*} \mid g_j)$
X_1	−0.791	0.377	<0.001	1.000	0.0925	−0.073	0.035
X_1, X_2	6.813	0.646	0.349	0.540	0.0682	0.465	0.044
X_1, X_3	−5.410	0.477	<0.001	0.797	0.0828	−0.448	0.040
X_1, X_4	−7.068	0.558	<0.001	0.591	0.0713	−0.504	0.040
X_1, X_2, X_3	6.142	1.292	0.154	0.114	0.0314	0.193	0.041
X_1, X_3, X_4	−8.079	0.575	0.233	0.616	0.0730	−0.590	0.042
$X_{1,} X_2, X_4$	6.868	1.494	0.128	0.106	0.0303	0.208	0.045
X_1, X_2, X_3, X_4	−2.050	6.315	0.136	0.005	0.0064	−0.013	0.041
$\hat{\bar{\beta}}_1$ and $\widehat{\mathrm{se}}\hat{\bar{\beta}}_1$:	2.040	6.376	$w_+(1) = 1.000$			0.079	0.342

β_2 rate of change with proportion herbaceous (15.2)

Predictors used	Unstandardized estimates					Standardized estimates	
	$\hat{\beta}_{2j}$	$\widehat{\mathrm{se}}\,(\hat{\beta}_{2j}\vert g_j)$	w_j	$\mathrm{VIF}_{2j}{}^{-1}$	Partial SD X_{2j}	$\hat{\beta}^*_{2j}$	$\widehat{\mathrm{se}}\,(\hat{\beta}^*_{2j}\vert g_j)$
X_1	0.000	0.000	<0.001			0.000	0.000
X_1, X_2	14.950	0.809	0.349	0.540	0.0549	0.821	0.044
X_1, X_3	0.000	0.000	<0.001			0.000	0.000
X_1, X_4	0.000	0.000	<0.001			0.000	0.000
X_1, X_2, X_3	14.223	1.456	0.154	0.139	0.0280	0.398	0.041
X_1, X_3, X_4	0.000	0.000	0.233			0.000	0.000
X_1, X_2, X_4	15.001	1.490	0.128	0.167	0.0306	0.459	0.046
X_1, X_2, X_3, X_4	6.066	6.326	0.136	0.007	0.0064	0.039	0.040
$\hat{\bar{\beta}}_2$ and $\widehat{\mathrm{se}}\hat{\bar{\beta}}_2$:	10.152	6.404	$w_+(2) = 0.767$			0.412	0.306

β_3 rate of change with proportion forest

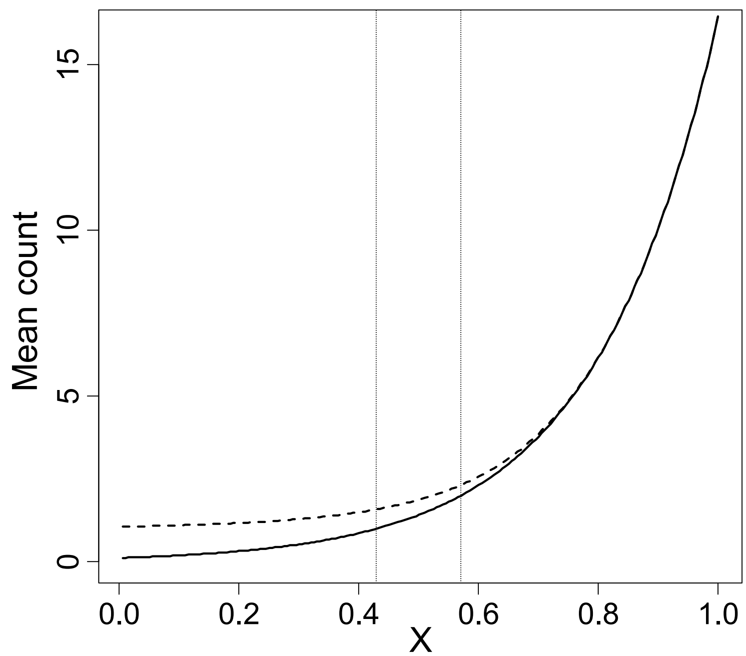
Predictors used	Unstandardized estimates					Standardized estimates	
	$\hat{\beta}_{3j}$	$\widehat{\mathrm{se}}\,(\hat{\beta}_{3j}\vert g_j)$	w_j	$\mathrm{VIF}_{3j}{}^{-1}$	Partial SD X_{3j}	$\hat{\beta}^*_{3j}$	$\widehat{\mathrm{se}}\,(\hat{\beta}^*_{3j}\vert g_j)$
X_1	0.000	0.000	<0.001			0.000	0.000
X_1, X_2	0.000	0.000	0.349			0.000	0.000
X_1, X_3	−21.350	1.401	<0.001	0.797	0.0447	−0.954	0.063
X_1, X_4	0.000	0.000	<0.001			0.000	0.000
X_1, X_2, X_3	−1.465	2.449	0.154	0.198	0.0224	−0.033	0.055
X_1, X_3, X_4	−15.129	1.493	0.233	0.613	0.0393	−0.595	0.059
X_1, X_2, X_4	0.000	0.000	0.128			0.000	0.000
X_1, X_2, X_3, X_4	−9.216	6.344	0.136	0.026	0.0082	−0.075	0.052

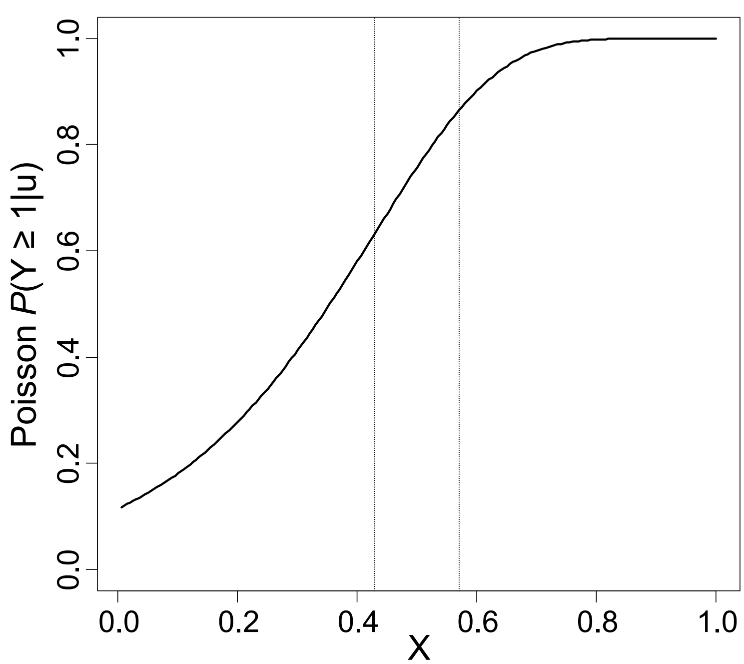
$\hat{\bar{\beta}}_3$ and $\widehat{\text{se}}\hat{\bar{\beta}}_3$:	−5.005	6.471	$w_+(3) = 0.522$			−0.154	0.210

β_4 rate of change with proportion alpine

Predictors used	Unstandardized estimates					Standardized estimates	
	$\hat{\beta}_{4j}$	$\widehat{\text{se}}\,(\hat{\beta}_{4j}\vert g_j)$	w_j	VIF_{4j}^{-1}	Partial SD X_{4j}	$\hat{\beta}^*_{4j}$	$\widehat{\text{se}}\,(\hat{\beta}^*_{4j}\vert g_j)$
X_1	0.000	0.000	<0.001			0.000	0.000
X_1, X_2	0.000	0.000	0.349			0.000	0.000
X_1, X_3	0.000	0.000	<0.001			0.000	0.000
X_1, X_4	−23.981	1.401	<0.001	0.591	0.0285	−0.683	0.040
X_1, X_2, X_3	0.000	0.000	0.154			0.000	0.000
X_1, X_3, X_4	−14.860	1.528	0.233	0.497	0.0262	−0.389	0.040
X_1, X_2, X_4	0.104	2.543	0.128	0.192	0.0163	0.003	0.078
X_1, X_2, X_3, X_4	−8.724	6.581	0.136	0.026	0.0060	−0.052	0.040
$\hat{\bar{\beta}}_4$ and $\widehat{\text{se}}\hat{\bar{\beta}}_4$:	−4.637	6.484	$w_+(4) = 0.498$			−0.098	0.142

FIG. B1. Top panel is exponential mean response curves for zero-truncated Poisson regression (dashed line) and corresponding Poisson regression (solid line) model generated by following R code: `x <- 1:200/200; mean.y <- exp(-2.1 + 4.9*x); ztp.mean.y <- exp(-2.1 + 4.9*x)/(1-exp(-exp(-2.1 + 4.9*x))); plot(x,ztp.mean.y,type="l", xlab="X", ylab="Mean count",ylim=c(0,16));par(new=T); plot(x,mean.y,type="l", xlab="X", ylab = "Mean count"); abline(v=c(0.4286, 0.570),lty=1).` Bottom panel is logistic response curve for associated probabilities of $Y \geq 1$ (occupancy) for Poisson regression given the means in the top panel generated by following R code: `ge1.mean.y <- 1 - ppois(0, mean.y); plot(x,ge1.mean.y,ylim=c(0,1),type="l",ylab="Poisson P(Y>=1|u)", xlab="X"); abline(v=c(0.4286,0.570),lty=1).` Dotted vertical lines in both panels indicate where a predicted mean count = 1.0 and its corresponding predicted probability of counts ≥1 = 0.632 (at x = 0.43) and where a predicted mean count = 2.0 and its corresponding predicted probability of counts ≥1 = 0.865 (at x = 0.57) given the Poisson distribution.


Mean count
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
X


Poisson P(Y ≥ 1|u)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
X