

An evaluation of prior influence on the predictive ability of Bayesian model averaging

Véronique St-Louis · Murray K. Clayton ·
Anna M. Pidgeon · Volker C. Radeloff

Received: 20 September 2009 / Accepted: 23 August 2011 / Published online: 23 September 2011
© Springer-Verlag (outside the USA) 2011

Abstract Model averaging is gaining popularity among ecologists for making inference and predictions. Methods for combining models include Bayesian model averaging (BMA) and Akaike's Information Criterion (AIC) model averaging. BMA can be implemented with different prior model weights, including the Kullback–Leibler prior associated with AIC model averaging, but it is unclear how the prior model weight affects model results in a predictive context. Here, we implemented BMA using the Bayesian Information Criterion (BIC) approximation to Bayes factors for building predictive models of bird abundance and occurrence in the Chihuahuan Desert of New Mexico. We examined how model predictive ability differed across four prior model weights, and how averaged coefficient estimates, standard errors and coefficients' posterior probabilities varied for 16 bird species. We also compared the predictive ability of BMA models to a best single-model

approach. Overall, Occam's prior of parsimony provided the best predictive models. In general, the Kullback–Leibler prior, however, favored complex models of lower predictive ability. BMA performed better than a best single-model approach independently of the prior model weight for 6 out of 16 species. For 6 other species, the choice of the prior model weight affected whether BMA was better than the best single-model approach. Our results demonstrate that parsimonious priors may be favorable over priors that favor complexity for making predictions. The approach we present has direct applications in ecology for better predicting patterns of species' abundance and occurrence.

Keywords Bayesian model averaging · BIC weights · Prior model weights · Predictive modeling · Chihuahuan Desert · Birds

Communicated by Wolf Mooij.

Electronic supplementary material The online version of this article (doi:10.1007/s00442-011-2118-6) contains supplementary material, which is available to authorized users.

V. St-Louis · A. M. Pidgeon · V. C. Radeloff
Department of Forest and Wildlife Ecology,
University of Wisconsin-Madison,
1630 Linden Dr., Madison, WI 53706, USA

V. St-Louis (✉)
Department of Fisheries, Wildlife, and Conservation Biology,
University of Minnesota, 1980 Folwell Ave, St. Paul,
MN 55108, USA
e-mail: vstlouis@umn.edu

M. K. Clayton
Department of Statistics, University of Wisconsin-Madison,
Madison, WI 53706, USA

Introduction

The desire to account for model uncertainty and increase predictive ability has motivated ecologists to depart from a single- to a multi-model approach to statistical inference (Burnham and Anderson 2002; Johnson and Omland 2004; Link and Barker 2006). The multi-model approach typically consists in deriving the coefficient estimates of the explanatory variables of interest by averaging the results from multiple models as opposed to drawing conclusions from a single model. The models used for the calculation are either chosen a priori based on prior knowledge of the phenomenon of interest or a posteriori using a user-defined criterion that will select only the best fitting models out of a pool of models. Burnham and Anderson (2001, 2002) provided a framework for implementing multi-model inference relatively easily based on the Akaike Information

Criterion (AIC). On the other hand, approaches such as Bayesian model averaging (BMA) have been used only sparingly by ecologists for making inference and prediction (Link and Barker 2006; Thomson et al. 2007) despite their popularity in statistics; Hoeting et al. (1999) provides a review of the origins of BMA. Ecologists' reluctance to use BMA may be due to the apparent complexity involved in implementing a full Bayesian approach, as opposed to the ease of implementation of AIC weights. However, the work of Link and Barker (2006) has shown that Bayesian Information Criterion (BIC) weights can provide a simple and flexible alternative to AIC model averaging. AIC weights are essentially equivalent to BIC weights with a different prior on the models. The drawback of AIC weights is that they may lead to a set of models that is more complex than desired (Link and Barker 2006). This suggests that the choice of the priors has strong effects on model selection. However, the predictive performance of BIC weights with different sets of priors is poorly understood.

Here, our goal was to compare the predictive ability of habitat models obtained using different priors, especially when using a BIC weights approximation to a full Bayesian model averaging. The priors can be viewed as model weights (hereafter prior model weights) that capture, independent of the data, the probability that a model is the best representation of reality among the pool of fitted models. These priors are subsequently used to calculate posterior model weights that depend on the data and that reflect, given the data, the probability that each model is the best representation of the data. The model weights ultimately determine how much a variable's coefficient estimate from a given model contributes to the final value of this variable's model averaged coefficient. The approach that we present is broadly applicable, has the advantage of being accessible to a wide range of scientists, and is straightforward to implement. Our study complements Link and Barker's (2006) work by specifically addressing the issue of the models' predictive ability, and tests an approach to Bayesian model averaging that is more extensive yet can still be easily implemented.

In general, model averaging approaches provide aggregate models with better predictive abilities than best single-model approaches (Raftery et al. 1997). For example, in a highly fragmented landscape in Australia, averaging over several models produces higher predictive ability than single "best" models for predicting the occurrence of 61 bird species (Thomson et al. 2007). Models addressing the effects of land-use and climate on the richness of seven groups of organisms in Europe were also more accurate when using a multi-model approach rather than a single-model one (Dormann et al. 2008). These examples

demonstrate the value of multi-model approaches for the specific purpose of building predictive models in ecology.

There are several ways of conducting model averaging, including AIC-based approaches and Bayesian approaches. Ecologists often use AIC weights to obtain coefficient estimates and variables' "relative importance weights" (Burnham and Anderson 2002). However, Link and Barker (2006) argue that most users of AIC weights are often unaware of the statistical assumptions underlying the use of AIC for model averaging (e.g., AIC weights favor models that have a higher number of parameters), and that other, more flexible approaches, such as BMA, may be preferable in some situations.

BMA uses so-called Bayes factors to construct model posterior probabilities (Eq. 1).

$$\Pr(M_i|data) = \frac{BF_{i,1}\pi_i}{\sum_j^R BF_{j,1}\pi_j}; \quad (1)$$

where $\Pr(M_i|data)$ is the posterior probability that M_i is true given that one of the $M_1, M_2, M_3, \dots, M_R$ is true (Link and Barker 2006), $BF_{i,1}$ are Bayes factors comparing models, π_i are the corresponding prior model weights and R is the total number of models. The models' posterior probabilities are used for calculating model averaged coefficients and standard errors as well as posterior probabilities for coefficients. The latter probabilities indicate the probability that a given coefficient is different from zero.

Estimating Bayes factors is complicated when a large number of predictors are involved, and it requires priors for the model parameters. The Bayesian Information Criterion (BIC), on the other hand, provides an approximation to the logarithm of the Bayes factors (Kass and Raftery 1995) such that:

$$BF_{i,j} \approx \exp(-(BIC_i - BIC_j)/2) \quad (2)$$

(Link and Barker 2006)

Using the latter approach and substituting the BIC approximation for the Bayes factors in Eq. 1, we obtain:

$$\Pr(M_i|data) \approx \frac{\exp(-BIC_i/2)\pi_i}{\sum_{j=1}^R \exp(-BIC_j/2)\pi_j} \quad (3)$$

The models' posterior probabilities obtained using Eq. 3 can thus be used as an alternative to AIC weights for conducting model averaging. Although a full Bayesian approach may be preferable (Link and Barker 2006; Link and Albers 2007), a BIC weights approximation can perform almost as well as the full BMA (Thomson et al. 2007), and is much easier to implement and thus a realistic approach for wider adoption in the ecological community.

There is a clear link between AIC weights and BIC posterior probabilities (Burnham and Anderson 2002).

Equation 3 is equivalent to the formulation of AIC weights (Burnham and Anderson 2002) when using a Kullback–Leibler model prior (Eq. 4).

$$\pi_i = \frac{\exp[k_i \log(N)/2 - k_i]}{\sum_{j=1}^R \exp[k_j \log(N)/2 - k_j]} \quad (4)$$

where k_i is the number of parameters in the model (including the intercept), N is the total number of observations, and R is the total number of models in the set.

The question is how the choice of prior model weights affects the predictive ability of models when implementing the approximation to BMA given in Eq. 3. It is not clear how the model averaged coefficients, standard errors, and posterior probabilities of the coefficients differ when using different prior model weights. Our main goal was to compare the ability of models built using four different prior model weights for predicting the abundance and occurrence of 16 bird species using the BIC approximation to obtain model averaged coefficients. We wanted to provide an understanding of the implications of selecting different types of prior model weights on the predictive ability of models obtained using BMA. Additionally, we wanted to compare BMA, using four different prior model weights, to a best single-model approach for making predictions.

Modeling approach

The approach we used was inspired by the “bicreg” and “bic.glm” functions available in the BMA package (Raftery et al. 2006) for R (R Development Core Team 2011). These functions use the BIC weights approximation (Eq. 3) and provide a simple alternative to the full implementation of BMA, which is beyond the scope of this paper. However, the existing functions assume a uniform prior on models. Here, we wrote a function similar to the “bicreg” and “bic.glm” functions of the BMA package but that allows comparing different prior model weights (Appendix C, D).

Our model averaging approach included three main steps. The first step was to fit all possible combinations of variables. Although this approach is often criticized as data dredging (Anderson and Burnham 2002), it is used properly here as a means for calculating posterior probabilities (Hoeting et al. 1999). The BMA package uses the leaps algorithm (Allen 1974) in a best subset analysis to reduce computing time, but our number of variables (see below) was low enough that we were able to explore the full set of all possible models. After fitting all combinations of models, we calculated the posterior probabilities of the models using Eq. 3, assuming uniform priors for the model

parameters. We then used Occam’s window (Madigan and Raftery 1994) to select a subset of models best supported by the data. Models not belonging to the set $\left\{ \frac{\max[\Pr(M_i|data)]}{\Pr(M_i|data)} \leq C \right\}$ were excluded. C is user-defined and set to a default of 20 in the BMA package (Raftery et al. 2006). The Occam’s window approach (Madigan and Raftery 1994) ensures that models with inconsequential support are discarded.

The second step was to recalculate the posterior probabilities for the models included in the subset so that when summed up, posterior probabilities added up to one. In the third and final step, posterior probabilities for the coefficients were obtained by combining posterior probabilities of only the models in which that particular variable occurs as in Eq. 5.

$$\Pr(\beta_l \neq 0) = \sum_{j=1}^R \Pr(M_j|data); \quad (5)$$

where N is the total number of models in which coefficient β_l occurs. The coefficient posterior probabilities obtained indicate the probability that each coefficient is different from zero. The posterior mean and variance of the coefficients are calculated as follows (Hoeting et al. 1999):

$$E[\beta_l|data] = \sum_{i=0}^R \hat{\beta}_{il} \Pr(M_i|data) \quad (6)$$

$$\text{Var}[\beta_l|data] = \sum_{i=0}^R \text{Var}[\beta_{il}|data, M_i] + \hat{\beta}_{il}^2 \Pr(M_i|data) - E[\beta_l|data]^2 \quad (7)$$

where $\hat{\beta}_{il}$ is the coefficient estimate for variable l in model i obtained by maximum-likelihood estimates, and R is the total number of models in the subset obtained from the Occam’s window criterion.

Case study

We examined the BIC weights approach to build predictive models of bird abundance and occurrence. Bird data were collected during the 1996 breeding season (May–June) at 42 12-point study grids randomly distributed across the seven habitat types covering the McGregor Range of Fort Bliss Army Reserve (New Mexico). Birds detected in a 10-min period within 150 m of each grid point were recorded four to five times during the breeding season. More details on the bird data are available in Pidgeon (2000) and Pidgeon et al. (2001). We summed the counts across the 12 points to get a plot-level measure of species abundance, and took the average of the two

highest visits to get a final measure of abundance at the plot level. Count data were square root transformed prior to the analysis.

We modeled a subset of 16 species occurring at more than 40% of the study sites. Eight common species were detected at more than 75% of the sites, and 8 occurred at 40–65% of the sites. Common species were modeled using linear models assuming a Gaussian error distribution while less common species were modeled using a generalized-linear logistic regression. The common species were Ash-throated Flycatcher (*Myiarchus cinerascens*) (ATFL, number of study plots present = 40), Black-throated Sparrow (*Amphispiza bilineata*) (BTSP, 37), Cactus Wren (*Campylorhynchus brunneicapillus*) (CACW, 34), Common Nighthawk (*Chordeiles minor*) (CONI, 37), Mourning Dove (*Zenaida macroura*) (MODO, 42), Northern Mockingbird (*Mimus polyglottos*) (NOMO, 41), Scott's Oriole (*Icterus parisorum*) (SCOR, 41), and Western Kingbird (*Tyrannus verticalis*) (WEKI, 39). Less common species were Brewers Sparrow (*Spizella breweri*) (BRSP, 24), Black-tailed Gnatcatcher (*Poliophtila melanura*) (BTGN, 20), Crissal Thrasher (*Toxostoma crissale*) (CRTH, 24), Eastern Meadowlark (*Sturnella magna*) (EAME, 26), Green-tailed Towhee (*Pipilo chlorurus*) (GTTO, 23), Pyrrhuloxia (*Cardinalis sinuatus*) (PYRR, 21), Scaled Quail (*Callipepla squamata*) (SCQU, 22), and Verdin (*Auriparus flaviceps*) (VERD, 18).

We quantified broad-scale habitat attributes in 1-km buffers around each point count using a classification from the Southwest ReGAP, created from Landsat Enhanced Thematic Mapper Plus (ETM+) imagery from 1999 to 2001 (Lowry et al. 2005). We first calculated the number of cover types in each buffer (patch richness; pr) and edge density of all classes (ed-allcl). Then, we reclassified the image into two classes, i.e., grasslands, and shrubland/woodland, to calculate the density of edges between these two cover types (ed-recl) and the proportion of shrubland/woodland cover (pshwo).

In addition to the broad-scale habitat attributes, we used an unclassified Landsat TM mosaic of June 1996 (path 33, rows 37 and 38) for quantifying within-habitat heterogeneity around each plot at an intermediate spatial scale. We used image texture analysis of the Normalized Difference Vegetation Index (NDVI, a measure of green biomass) to quantify the degree of variability in pixel values in a 9×9 window, an area roughly equal to the extent of a 150-m radius point count. Image texture of NDVI is useful for discriminating habitat types in this ecosystem (St-Louis et al. 2009). We quantified first-order mean (mean) and coefficient of variation (cv), as well as second-order angular second moment (asm), contrast (con), and correlation (corr) (Haralick et al. 1973). We extracted elevation (elev) at each point count from a 10-m digital elevation model. Broad- and

intermediate-scale habitat data, as well as elevation, were averaged across the 12 points for obtaining plot-level measures of habitat that matched the bird data.

To construct the habitat models, we implemented model averaging using BIC weights with four different prior model weights (Link and Barker 2006): (1) Uniform prior; $\pi_i = \frac{1}{R}$, where R is the total number of models considered, (2) Occam's prior of parsimony; $\pi_i = \exp(-k)$, where k is the number of parameters, (3) Complexity weights favoring complex models; $\pi_i = \exp(k)$, and (4) the Kullback–Leibler prior (Eq. 4), which is equivalent to AIC weights (Link and Barker 2006).

The total number of models before applying the Occam's window criterion was 1,023 (all possible combinations of 10 variables). We chose a constant $C = 20$ for implementing Occam's window, which has been shown to provide good predictive performance (Raftery et al. 1996). For each combination of the 16 species and the four priors, we used the BMA approach outlined above to obtain model-averaged coefficients and standard error estimates, as well as coefficient posterior probabilities. To compare with more traditional model performance metrics, we used an F test for calculating the overall significance of the models that were included in the parsimonious subset and calculated the adjusted coefficient of determination (R_{adj}^2).

We evaluated the predictive ability of the Bayesian averaged models using leave-one-out cross-validation. We iteratively re-fitted the BMA for each observation i to obtain new coefficient estimates based on the reduced data set (i.e., $N - 1$ observations). These coefficients were then used to predict the value of observation i . We calculated the Predicted Residual Sum of Squares (PRESS) statistics (Allen 1974) for evaluating model predictive ability as follows:

$$\text{PRESS} = \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (8)$$

where y_i is the value of the i th observation, \hat{y}_i it the predicted value of the i th observation using the reduced model, and N = the number of observations (here $N = 42$). For the logistic regression models that were applied to the eight least common species, \hat{y}_i corresponds to the predicted probability of occurrence of observation i , calculated as:

$$\hat{y}_i = \frac{\exp(\hat{m}_i)}{1 + \exp(\hat{m}_i)} \quad (9)$$

where \hat{m}_i is the predicted value calculated from the model averaged coefficients for observation i . Lower values of PRESS indicate better predictive models.

To compare the predictive ability of a BMA to a best single-model approach, we calculated the PRESS statistics

from predictions obtained using best models only. For each set of $N - 1$ observations, we fitted all possible combinations of models and found the best fitting model using BIC. This model was then used to predict the abundance, or probability of occurrence value at observation i . This approach therefore matched the approach that was used to calculate the PRESS statistics of the BMA models where BMA was re-fitted for each set of $N - 1$ observations. The predicted values were then used to calculate PRESS using the same formula as for BMA.

Results

From the list of 1,023 possible combinations of variables, the number of best supported models as defined by Occam’s window was generally small, with an average across all species of 56 (30–120 models depending on the species), 16 (11–30), 323 (54–73) and 276 (55–549) models for Uniform, Occam’s, Complexity, and Kullback–Leibler priors, respectively. The number of parameters of these best supported models varied across prior model weights, with models generally containing no more than five to six parameters for Occam’s and Uniform priors, and larger models of up to ten parameters (i.e., full model) for Complexity and Kullback–Leibler priors (Table 1; Appendix A).

Predictions obtained from BMA were better than a best single-model approach for 12 out of 16 species for some or all of the prior model weights (Table 2; Fig. 1). Using the Scott’s Oriole and the Crissal Thrasher as examples, Fig. 1 shows that for these species the absolute error of the predictions obtained using the BMA approach with the Occam’s prior model weight versus the best single-model approach is lower at most of the 42 study sites. Due to very low significance of models for the Mourning Dove we feel

it is not useful to consider that species, and restrict further consideration to the remaining fifteen species. For 6 species out of the remaining 15, BMA was better than a best single-model approach for all prior model weights. However, for species such as Black-throated Sparrow, Common Nighthawk, Crissal Thrasher, Verdin, and Western Kingbird, the predictive ability of models built using BMA was better than the single-model approach when implemented with only some of the prior model weights. For Western Kingbird, for example, the BMA approach was better than the best single-model approach with the Occam’s prior only. For other species such as the Black-tailed Gnatcatcher, the Common Nighthawk, and the Verdin, both the Uniform and Occam’s prior led to better predictive ability than the best-model approach, but the Kullback–Leibler and Complexity priors were inferior or equivalent. The PRESS statistic varied among prior model weights (Table 2). Occam’s priors resulted in the lowest PRESS statistics for 10 out of 15 species. Among the four prior model weights, the PRESS values of the Uniform prior were the second-lowest for most species. The modeling strategy that puts a higher weight on more complex models (i.e., Complexity or Kullback–Leibler priors) did not lead to lower PRESS statistics, except in a few cases (e.g., Ash-throated Flycatcher and Black-throated Sparrow). The Kullback–Leibler model provided overall the second-, or third-best predictions.

The Uniform and Occam’s priors led to smaller posterior probabilities for the coefficients for most habitat variables compared to the Complexity or Kullback–Leibler priors (Table 3, Appendix B). The Occam’s prior led to the smallest coefficient posterior probabilities, with only one or two, if any, over 50% for each species. On the other hand, several variables had coefficients’ posterior probabilities larger than 50% when using the Complexity and Kullback–Leibler priors. The model averaged coefficients and

Table 1 Examples of overall model fit statistics for the list of models (M) best supported by the data under the Occam’s window criterion of 20 for each prior

Species	Prior	Size	BIC	$R^2_{adj.}$	F statistic	p value	M
CRTH	Uniform	1, 4	51, 57	na	4.8, 13.4	0, 0.002	52
	Occam	1, 3	51, 55	na	6.3, 13.4	0, 0.001	11
	Comp.	1, 8	51, 69	na	2.8, 13.4	0, 0.005	551
	KL	1, 7	51, 66	na	3.1, 13.4	0, 0.003	410
SCOR	Uniform	2, 6	81, 86	47.3, 57.7	9.7, 19.7	0, 0	72
	Occam	2, 5	81, 87	46.3, 57.6	11.7, 19.7	0, 0	30
	Comp.	3, 9	81, 96	51.2, 57.7	6.3, 16.5	0, 0	204
	KL	3, 9	81, 95	51.2, 57.7	6.5, 16.5	0, 0	193

Results are presented for a species modeled using logistic regression (CRTH), and a species modeled using linear models (SCOR). The values represent range in model size (i.e., number of parameters), BIC, $R^2_{adj.}$, and F statistics and associated p value for the M models. Refer to “Case study” for species’ acronym description. A table for all species is provided in Appendix A

CRTH Crissal Thrasher, SCOR Scott’s Oriole

Table 2 Cross-validated PRESS statistic obtained using the multi-model approach with four different model priors and the best single-model approach

ATFL Ash-throated Flycatcher, *BRSP* Brewer's Sparrow, *BTGN* Black-tailed Gnatcatcher, *BTSP* Black-throated Sparrow, *CACW* Cactus Wren, *CONI* Common Nighthawk, *CRTH* Crissal Thrasher, *EAME* Eastern Meadowlark, *GTTO* Green-tailed Towhee, *MODO* Mourning Dove, *NOMO* Northern Mockingbird, *PYRR* Pyrrhuloxia, *SCOR* Scott's Oriole, *SCQU* Scaled Quail, *VERD* Verdin, *WEKI* Western Kingbird

Species	Model type	Prior model weights					Best model
		Uniform	Occam	Complex	KL		
ATFL	Gaussian	25.2	22.6	22.5	23.2	33.7	
BRSP	Logistic	11.0	10.6	11.9	11.7	10.3	
BTGN	Logistic	8.9	8.8	9.3	9.3	9.3	
BTSP	Gaussian	37.8	39.3	35.6	35.8	36.2	
CACW	Gaussian	22.1	21.0	21.8	21.8	27.7	
CONI	Gaussian	18.7	18.2	20.4	20.1	19.8	
CRTH	Logistic	8.2	8.1	9.3	9.1	8.8	
EAME	Logistic	6.8	6.6	7.2	7.1	5.8	
GTTO	Logistic	9.0	9.1	10.1	9.8	8.3	
MODO	Gaussian	43.0	42.3	47.5	46.3	40.7	
NOMO	Gaussian	33.4	32.3	34.5	34.5	38.1	
PYRR	Logistic	5.2	3.9	6.8	6.6	9.0	
SCOR	Gaussian	23.8	18.8	26.7	26.4	28.0	
SCQU	Logistic	7.1	6.7	7.8	7.8	11.5	
VERD	Logistic	9.2	8.5	10.6	10.6	10.1	
WEKI	Gaussian	35.2	33.0	37.7	37.2	33.4	

standard errors also varied among prior model weights, but the values were similar for variables with high posterior probabilities.

Discussion

The purpose of our study was to compare the predictive ability of models built using different prior model weights, especially when using a simple, BIC weights approximation of a full Bayesian model averaging. The take-home messages derived from our results are twofold. First, a simple implementation of BIC weights and the associated prior has the ability to produce better predictive results than using AIC weights (i.e., the Kullback–Leibler prior) (Burnham and Anderson 2001, 2002) or using a best single-model approach. Second, the BIC weights approach can be implemented relatively easily in any statistical package. Furthermore, one can achieve better predictive ability with only minor modifications of the default prior model weights. Our results complement and extend the work of Link and Barker (2006) by showing that their findings (i.e., more conservative priors may be best in some situations) hold in a predictive modeling framework, and by providing a simple methodology for implementing an approximation of the full BMA.

Overall, the predictive ability of the models built using BMA was superior to that of best single-models. Using BMA for making predictions therefore never hurts and frequently helps obtain better predictive ability. This is supported by recent examples in the literature (e.g., Dormann et al. 2008; Thomson et al. 2007). What our results show, however, is that whether or not BMA is better than a best single-model approach for making predictions can depend on the prior model weight. For six species, the PRESS statistic values we obtained for the best single-

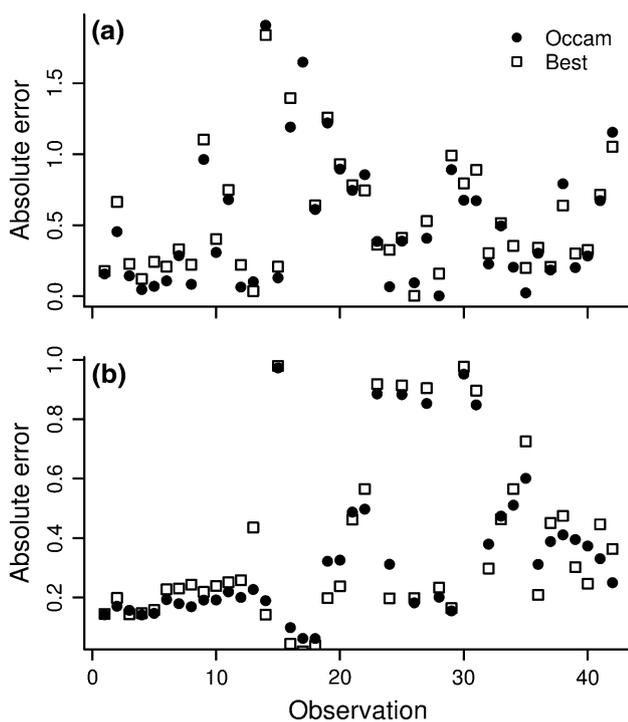


Fig. 1 Absolute prediction error of **a** Scott's Oriole's abundance (*sqrt*) and **b** Crissal Thrasher's probability of occurrence across the 42 sites for models generated using the BMA approach with Occam's prior model weight (*Occam*) and the best model approach (*Best*)

Table 3 Example of coefficients' posterior probabilities [$P(\neq 0)$] and model averaged coefficients and standard errors obtained from averaging between 11 and 410 models (Table 1) for CRTH for the four model priors

Variables	Uniform		Occam's		Complexity		Kullback–Leibler	
	P ($\neq 0$)	Mean (SE)	P ($\neq 0$)	Mean (SE)	P ($\neq 0$)	Mean (SE)	P ($\neq 0$)	Mean (SE)
Asm	11	0.05 (0.27)	2	0 (0.07)	39	0.28 (0.66)	34	0.24 (0.59)
con	11	−0.05 (0.65)	4	0.02 (0.16)	37	−0.39 (2.2)	32	−0.33 (1.89)
corr	18	0.63 (2.03)	5	0.1 (0.85)	49	1.88 (3.29)	45	1.69 (3.14)
cv	25	0.38 (0.98)	9	0.1 (0.45)	57	1.18 (1.73)	52	1.04 (1.63)
Mean	16	0.13 (0.54)	6	0.04 (0.22)	44	0.68 (1.44)	39	0.59 (1.3)
elev	90	−1.59 (0.83)	98	−1.61 (0.63)	81	−1.56 (1.28)	82	−1.57 (1.19)
ed_rcl	35	0.36 (0.62)	20	0.19 (0.46)	52	0.53 (0.8)	49	0.51 (0.77)
ed_allcl	25	0.22 (0.49)	13	0.1 (0.31)	44	0.43 (0.84)	40	0.39 (0.78)
pshwo	17	0.13 (0.39)	6	0.04 (0.21)	43	0.21 (0.76)	38	0.2 (0.68)
Pr	9	0 (0.24)	5	0.03 (0.15)	35	−0.23 (0.7)	29	−0.17 (0.61)

Bold indicates the only case where the coefficient would be considered significant based on a 95% confidence interval. Results for all 16 species are presented in Appendix B. Refer to “[Case study](#)” for species' acronym description

model approach were solidly situated among those obtained for the BMA models built using different prior model weights. In one case, more complex prior model weights led to models with higher predictive ability than a best single-model approach, while in three other cases, it was the exact opposite (i.e., parsimonious priors were best). This suggests that more research is needed to fully understand how much of an improvement BMA provides over a best single-model approach when different model prior model weights are used.

Among the prior model weights we studied, the ones that favored simplicity generally led to higher predictive ability. Occam's prior most often provided the best predictive models of bird abundance and occurrence. This is supported by the findings of Thomson et al. (2007), where prior model weights that favor simpler models had better predictive ability. As eluded to in Jefferys and Berger (1991), a smaller set of hypotheses may provide better predictions because there is less posterior weight put on extraneous models. This fits our findings. Generally, we obtained better predictions using the Occam's prior model weight as opposed to less parsimonious priors, and in the logic of Jefferys and Berger (1991), this happens because a smaller set of models with only a few parameters that capture the response variable (here bird abundance or occurrence) well is sufficient, and more likely to be accurate, than a larger set of models that contains many parameters with a diluted posterior. Interestingly, AIC weights are supposed to be a balance between parsimony and complexity (Burnham and Anderson 2004). In contrast, our results suggest that AIC weights favor complexity over parsimony, which corresponds to Link and Barker's (2006) finding. At least it is valuable to researchers to be aware of a broad spectrum of priors, and analytical methods, so that these might be better matched to the goals

of a given study. Given similar predictive ability, we recommend using prior model weights that favor parsimony over complexity, especially if the sample size is small. Methods that identify fewer, but most ecologically relevant, variables are valuable, since measuring many variables, especially in field studies, is costly. With that in mind, we believe that tools such as the BMA package could be greatly enhanced by allowing the users to modify the default prior model weight. Priors that favor parsimony such as the Occam's prior have the advantage for conservation science and practice of (1) emphasizing parameters that may be more biologically relevant, (2) favoring a conservative use of covariates, (3) saving resources and computing time, and (4) being a better modeling strategy when faced with small sample sizes.

We demonstrated a method for rapidly obtaining model-averaged coefficients that can be used for making predictions. The Occam's window-based approach (Madigan and Raftery 1994), available in the BMA package (Raftery et al. 2006) for R (R Development Core Team 2011) and in the function that we provide in Appendix D, has the advantage of being fast and easy to implement. While not as thorough as conducting a full BMA with priors on parameters and models (Link and Barker 2006), we believe that it is a useful alternative to the commonly used AIC model averaging for building predictive models in ecology, based on the leave-one-out PRESS statistics we obtained.

We have chosen to use the leave-one-out PRESS statistics over other model validation methods because of our small sample size. Other methods, such as bootstrap, k-fold cross-validation, or a validation of the models against an independent dataset, would be valuable for studies conducted over larger sample sizes and where independent data are available.

In summary, our results show that the BIC weights approach is a good, more flexible alternative to AIC model averaging with its default prior model weight and improves the predictive ability of the models when using more parsimonious prior model weights. Link and Barker (2006) showed the tendency of the Kullback–Leibler prior to favor more complex models. Our results extend this, and clearly show that the AIC weights favoring more complex models do not increase the predictive ability of the averaged models.

Acknowledgments We would like to thank the subject editor, Dr. Wolf M. Mooij, and two anonymous reviewers for valuable comments on the manuscript. We would also like to thank Dr. James D. Forester for his help with the analysis and R code. We thank all field workers who contributed to the data collection in 1996. Support for this research was provided by the U.S. Strategic Environmental Research and Development Program, Legacy Resource Management Program, the Fort Bliss Environmental Division-Directorate of Public Works, USGS BRD Texas Cooperative Fish and Wildlife Research Unit, USGS BRD Wisconsin Cooperative Wildlife Research Unit, and the Department of Forest and Wildlife Ecology, University of Wisconsin-Madison.

References

- Allen DM (1974) The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 16:125–127
- Anderson DR, Burnham KP (2002) Avoiding pitfalls when using information-theoretic methods. *J Wildl Manag* 66:912–918
- Burnham KP, Anderson DR (2001) Kullback–Leibler information as a basis for strong inference in ecological studies. *Wildl Res* 28:111–119
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York
- Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–304
- Dormann CF, Schweiger O, Arens P, Augenstein I, Aviron S, Bailey D, Baudry J, Billeter R, Bugter R, Bukacek R, Burel F, Cerny M, De Cock R, De Blust G, DeFilippi R, Diekötter T, Dirksen J, Durka W, Edwards PJ, Frenzel M, Hamersky R, Hendrickx F, Herzog F, Klotz S, Koolstra B, Lausch A, Le Coeur D, Liira J, Maelfait JP, Opdam P, Roubalova M, Schermann-Legionnet A, Schermann N, Schmidt T, Smulders MJM, Speelmans M, Simova P, Verboom J, van Wingerden W, Zobel M (2008) Prediction uncertainty of environmental change effects on temperate European biodiversity. *Ecol Lett* 11:235–244
- Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC* 3:610–621
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Stat Sci* 14:382–417
- Jefferys WH, Berger JO (1991) Sharpening Ockham's razor on a Bayesian stop. Technical Report 91-44C, Department of Statistics, Purdue University
- Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends Ecol Evol* 19:101–108
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90:773–795
- Link WA, Albers PH (2007) Bayesian multimodel inference for dose-response studies. *Environ Toxicol Chem* 26:1867–1872
- Link WA, Barker RJ (2006) Model weights and the foundations of multimodel inference. *Ecology* 87:2626–2635
- Lowry JHJ, Ramsey RD, Bradford D, Comer P, Falzarano S, Kepner W, Kirby J, Langs L, Prior-Magee J, Manis G, O'Brien L, Sajwaj T, Thomas CD, Rieth W, Schrader S, Thompson B, Wallace C, Waller DM, Wolk B (2005) Southwest Regional Gap Analysis Project: final report on land cover mapping methods. Logan, Utah
- Madigan D, Raftery AE (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *J Am Stat Assoc* 89:1535–1546
- Pidgeon AM (2000) Avian abundance and productivity at the landscape scale in the Northern Chihuahuan Desert. University of Wisconsin-Madison, Madison
- Pidgeon AM, Mathews NE, Benoit R, Nodheim EV (2001) Response of avian communities to historic habitat change in the Northern Chihuahuan Desert. *Conserv Biol* 15:1772–1788
- Raftery AE, Hoeting JA, Volinsky CT, IS Painter, Yeung KY (2006). BMA: Bayesian Model Averaging. R package version 3.03. <http://www.r-project.org>, <http://www.research.att.com/~volinsky/bma.html>
- Raftery AE, Madigan D, Volinsky CT (1996) Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). In: Bernardo J, Berger J, Dawid A, Smith A (eds) *Bayesian Statistics 5*. Oxford University Press, USA, pp 323–349
- Raftery AE, Madigan D, Hoeting JA (1997) Bayesian model averaging for linear regression models. *J Am Stat Assoc* 92:179–191
- St-Louis V, Pidgeon AM, Clayton MK, Locke BA, Bash D, Radeloff VC (2009) Satellite image texture and vegetation indices predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography* 32:468–480
- R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org>
- Thomson JR, Mac Nally R, Fleishman E, Horrocks G (2007) Predicting bird species distributions in reconstructed landscapes. *Conserv Biol* 21:752–766