

Using Bayesian Model Averaging to Better Represent Uncertainty in Predictions of Ecological Models

B. Wintle^a, M.A. McCarthy^{a,b} and R.P. Kavanagh^c

^a*School of Botany, University of Melbourne, Parkville VIC, Australia (b.wintle1@pgrad.unimelb.edu.au)*

^b*ARCUE, Royal Botanic Gardens Melbourne, VIC, Australia(mamcca@unimelb.edu.au)*

^c*State Forests of NSW, NSW, Australia (rodk@sf.nsw.gov.au)*

Abstract: Model selection uncertainty arises when one model is chosen over plausible alternatives. Representations of prediction uncertainty routinely incorporate uncertainty about parameter estimates contingent on the choice of a single best model that is used to represent truth. However, classical prediction intervals routinely fail to incorporate uncertainty about the choice of that model, and as a result are often optimistically tight. Optimistically tight prediction intervals can lead to overconfident decisions that do not incorporate sufficient hedging against uncertainty. Bayesian Model Averaging (BMA) provides a simple means of incorporating model selection uncertainty into statistical inference and prediction. This paper gives two examples of the application of BMA in activities fundamental to conservation biology; predicting the spatial distribution of wildlife species, and assessing the risk of decline in populations of threatened species.

Keywords: Bayesian model averaging; Habitat model; Logistic regression; Prediction uncertainty; Population viability analysis.

1. INTRODUCTION

Often we seek to make predictions about unknowns (such as the probability of occurrence or the future population size of a threatened species) based on a set of things that are known (such as aspects of the environment, or the biology of the species). We seldom know the true relationship between what we hope to predict and the information at hand, so we rely on a model to formalise our judgment. A model will generally comprise two parts. Firstly, the model will have a structure that is defined by the choice of the functional relationships, variables, transformations and interactions. Secondly, a model will have a set of parameter estimates that are specific to a given model structure. For example, Lindenmayer et al. [1990] modelled the probability of occurrence of Greater Gliders (y) as a function of forest age (a) and the number of trees (greater than 0.5m diameter) with cavities in a given 3 ha cell (n)

$$\text{logit}(y) = \ln[y/(1-y)] = -0.993 + 1.106a + 0.554n. \quad (1)$$

The assumptions made by the authors included:

- only the variables a and n influence the presence or absence of the species;
- a logit link function rather than other alternatives was suitable; and

- there was additive linearity in the predictor variables resulting in no transformation of, or interaction between the predictor variables.

Clearly, there are numerous alternative models that might result from the combination of possible assumptions listed above. In practice, the structural features of the model are often determined by some data-driven search over the possible models that could be chosen to represent our judgment about the relationship of interest. It is usual to make predictions contingent on the 'best' model being 'correct', considering only the uncertainty about the parameters in the computation of prediction intervals. This assumption is tenuous given that errors arising from uncertainty about the structure of the model are likely to be far worse than those arising from other sources [Chatfield, 1995].

In ecology, there is often ambiguity about the processes involved in determining the true value of the response. In many instances, there are a number of possible model structures that fit the observed data almost as well as the chosen 'best' structure. Hence, there is considerable uncertainty about which model is in fact the best. If plausible alternative models result in predictions that are very different from those of the chosen 'best' model, there is the risk of ignoring such alternatives. A more conservative approach might

be to think of any given model as a sample from an infinite set of competing models, then use data to weight or assign a degree of belief to the competing models. Below, we highlight how model selection uncertainty is naturally incorporated within a Bayesian framework using Bayesian Model Averaging (BMA), [Draper, 1995].

In some instances a small set of competing models may represent well-defined schools of thought or competing hypotheses regarding a biological process (section 3). These models might be adopted as the set of competing models, keeping in mind that each competing hypothesis might be plausibly represented by more than one model. In these truly 'discrete' situations, the only impediment to a simple BMA analysis is the identification of a coherent scheme for assigning posterior model probabilities to candidate models.

In many situations, such as multiple regression for the construction of habitat models (section 2), the number of candidate variables can be very large and the number of possible model structures becomes enormous. In this case, we utilise an approach developed by Madigan and Raftery [1994] and refined by Raftery et al. [1996] called Occam's window by which we can identify a subset of parsimonious candidate models that are supported by the data. This method firstly excludes models where the ratio of posterior model probabilities is strongly in favour of one model, such that the favoured model is greater than twenty times as likely as the rejected model. Secondly, where a smaller (nested) model is given greater support by the data (has a higher posterior probability) than a larger model, the larger model is excluded. Code entitled BIC.GLM (www.research.att.com/~volinsky/bma.html) enables execution of this method in Splus, and is discussed in more detail in Hoeting et al. [1999] and Raftery et al. [1996].

The following examples are presented to illustrate the simplicity with which BMA methods can be applied in standard ecological modelling situations.

2. BMA FOR SPATIAL PREDICTION

This example demonstrates how the single-best-model approach to habitat modelling can be extended to the BMA approach, enabling the incorporation of model selection uncertainty. We construct habitat models of the Greater Glider (GG) using the BMA and single-best approaches. Using new data, we then test the predictive accuracy and coverage of the 95% prediction intervals for the models that were developed.

2.1. Background

The GG is an arboreal, folivorous, gliding marsupial whose distribution extends from central western Victoria to just north of the tropic of Capricorn [McKay, 1995]. Being an obligate hollow nester, it is dependent on old growth forest characteristics [Lindenmayer et al., 1990]. Cork and Catling [1996] categorise the bulk of the studies on arboreal mammals into those pursuing the hypothesis that the foliar nutrient status of the forest is the prime determinant of habitat quality for arboreal mammals [Braithwaite, 1983], and those that put equal or greater emphasis on variables related to structural characteristics of the forest [e.g., Davey, 1989 and Lindenmayer, 1990]. Cork et al. [1994] provide evidence of a hierarchical process within which forest structural attributes act as a determinant of GG habitat above a certain threshold of foliar nutrient. This hypothesis was supported by the work of Pausas et al. [1995], and provides us with a basis for determining a set of candidate predictors for the probability of occurrence of the GG.

2.2. The Competing Models

Surveys of the presence of owls and gliders were conducted by Kavanagh and Bamkin [1995] and Kavanagh [1997] in the Eden region of southern NSW in 1992 and 1994 respectively. Data from a total of 187 of the 1992 survey sites were used for the test data set and data from 219 of the 1994 sites were used to develop the models.

We compiled a list of candidate predictors based on the prior habitat analyses discussed above. For the purpose of model building and testing, it was necessary that predictors were available as mapped digital layers in the Eden region that could be stored and handled in ArcView [ESRI, 1998]. The variables considered as candidate predictors were foliar nutrient index, area of old growth within 1 km, the area of old growth where the foliar nutrient index was greater than 4, terrain position, amounts of wet forest and rainforest within 300 m, mean annual rainfall and topographic wetness index. Some potentially useful variables such as stand basal area and time since logging or fires were not available for all survey sites and consequently could not be included as candidate predictors.

Models were fitted to the data using Splus after removing correlated variables, and BIC values were obtained using the BIC.GLM code of Raftery [1996]. Three models were identified as the plausible set of competing models using the BMA approach outlined above

$\text{logit}(y) = 1.129 - 0.051T + 0.032F + 0.004W,$
 $\text{logit}(y) = 3.047 - 0.045T + 0.031F + 0.078R,$
 $\text{logit}(y) = 3.47 - 0.047T + 0.031F,$

where y is the probability of presence of a GG, T is mean annual temperature, F is the proportion of old-growth forest within 1km where foliar nutrient index is greater than 4, W is the wetness index score for the 100m grid cell, and R is the proportion of rainforest within a 300m radius. The posterior probabilities of the three models, computed as a ratio of model BIC values, were 0.49, 0.28 and 0.23 respectively.

We made predictions for 187 of Kavanagh's 1992 survey sites using both the single-best model (model 1 above) and the BMA prediction from all three competing models. Both sets of predictions were then compared to the observations from the 1992 survey using Mann-Whitney U statistics [Pearce and Ferrier, 2000]. Both the single-best and model averaged predictions performed reasonably well with Mann-Whitney statistics of 0.81 and 0.82 respectively, where a score of 1 implies perfect discrimination and a score of 0.5 implies predictive discrimination that is no better than a random guess.

We calculated the HL goodness of fit statistic (\hat{C}), [Hosmer and Lemeshow, 1989] for both the single-best and BMA predictions at the independent 1992 sites. Single-best and BMA approaches returned \hat{C} values of 5.40 and 7.78 respectively, showing no substantial evidence of a lack of fit between the models and the new data ($P=0.72$ and 0.46).

2.3. Predictive Coverage of Single-Best and BMA Prediction Intervals

Computation of point-wise prediction intervals for a single model requires the calculation of the large sample asymptotic standard error (ASE) [Agresti, 1996]

$$ASE(\hat{p}) = \sqrt{\text{Var}(\hat{\alpha}) + x^2 \text{Var}(\hat{\beta}) + 2x \text{Cov}(\hat{\alpha}, \hat{\beta})}, \quad (2)$$

where $\hat{\alpha}$ represents the model intercept parameter and $\hat{\beta}$ represents a vector of estimated model coefficients. A pointwise prediction interval can be computed as

$$\hat{p} \pm t_{\alpha} ASE. \quad (3)$$

Computation of prediction intervals for the BMA predictions requires the integration of the intervals for each of the competing models. Buckland et al. [1997] describe a method that enables the computation of pointwise ASEs

$$ASE(\hat{p}_a) = \sum_{i=1}^R w_i \sqrt{\text{var}(\hat{p}_i) + (\hat{p}_i - \hat{p}_a)^2}, \quad (4)$$

where w_i is the weight assigned to model i , which is the posterior probability of model i in our example, $\text{var}(\hat{p}_i)$ is the variance of the linear predictor for model i derived using (2) above, and $\hat{p}_i - \hat{p}_a$ is the difference between the prediction from model i and the weighted average prediction from all R models. The 95% confidence intervals for predictions to Kavanagh's 1992 survey sites based on the single-best model were computed using (2) and (3). The 95% confidence intervals for model averaged predictions at the same sites were computed using (3) and (4).

Predictive coverage measures the performance of prediction intervals according to how often the intervals for a particular level of prediction contain the observed proportion of occurrences [Hosmer and Lemeshow, 1995]. We measured the predictive coverage of intervals using an approach analogous to that of the HL-test for goodness of fit. Observations with similar predictions were grouped for both the single-best and BMA predictions. Each group was then assigned an upper and lower 95% interval that was the average of the bounds for each prediction within the group. The number of times that the observed proportions of presence fell within the 95% interval for probability of presence was assessed for both single-best and BMA model predictions. Observed proportions of presence fell within the BMA prediction intervals 90% of the time, while the single-best intervals bounded the observed proportion of presence 70% of the time. The BMA interval was closer to the nominal (95%) coverage than the single best interval.

2.4. Interpretation

In contrast to Draper [1995] and Hoeting et al. [1999], the single-best model made similar predictions in terms of discrimination and calibration to those of the model average. This may be due in part to the discrete set of competing models used to derive the averaged predictions or that the data used to test predictions was collected in the same general location and using the same methods as the model building set. However, the usefulness of predictions can be thought of both in terms of the proportion of times that the mean prediction is correct, and how well we estimated our uncertainty about the future observations. In the example presented above, the prediction intervals computed with the BMA approach exhibited closer to nominal coverage than those derived from the single-best model approach. This indicates that even though we were considering a

small set of competing models, there was still considerable uncertainty about which model should be used.

3. BMA FOR POPULATION MODELLING

Population models are used in conservation biology to predict risks of extinction [Burgman et al., 1993]. Approaches to developing these models involve choosing an appropriate model and estimating the parameters for this model. There has been some consideration of the uncertainty due to parameter estimates [Taylor, 1995; Ludwig, 1996a], and uncertainty about the choice of the appropriate model is acknowledged [e.g., Ludwig, 1996a; McCarthy, 1996]. However, the absence of a recognized method to integrate the predictions of different models means that a single model is usually chosen for prediction, despite the fact that there is considerable uncertainty about whether the choice is correct. Moreover, in the uncommon cases where prediction intervals are provided for PVA models, uncertainty of model choice has not been incorporated [Ludwig, 1996a].

3.1. Methods

In this example, we examine different models for predicting risks of population decline of the mountain pygmy possum (*Burramys parvus*), a threatened species confined to alpine boulder fields in southeastern Australia. The species breeds at one year of age, so a model without age structure was used. Three different forms of density dependence were considered, based on those presented by Burgman et al. [1993, pp. 85-87] for unstructured populations. These took the form $N_{t+1} = g_t N_t$, where N_t is the population size at time t , and g_t is the annual expected growth rate. The models were the Ricker model

$$g_t = e^{r-hN_t}, \quad (5)$$

the Hassell and May model

$$g_t = R/(1+aN_t)^c, \quad (6)$$

and the Maynard-Smith and Slatkin model

$$g_t = R/(1+(pN_t)^q). \quad (7)$$

Environmental stochasticity was modelled by drawing the population growth rate in each year from a gamma distribution with coefficient of variation v , and mean g_t . Demographic stochasticity was modelled by drawing the number of individuals in the next year, given a particular expected growth rate, from a Poisson distribution. The net result of this is that the number of individuals in the next year is drawn from a negative binomial distribution [Ludwig, 1996b].

Data used to fit the models were obtained from McCarthy and Broome [2000]. The likelihood for each annual transition was obtained from the negative binomial distribution [Ludwig, 1996b], and the overall likelihood was obtained as the product of these values. For the data given by McCarthy and Broome [2000], there are four different populations. It was assumed that the maximum growth rate and shape of density dependence was the same for all populations and that the scale parameter (b in (5), a in (6), p in (7)) varied between each population depending on the equilibrium population size (such that $g_t=1$).

In the example presented here, predictions were made for an equilibrium population size of 30, so these scale parameters were fixed for given values of the other parameters. Thus, the unknown parameters were v , and also r in (5), R and c in (6), and R and q in (7). The prior distributions for the unknown parameters were determined from the biology of the species. Given the long-term (> several thousand year) persistence of the species in isolated populations, it was assumed that the maximum population growth rate was greater than 1. Given that each female produces one litter of up to 4 offspring each year, and assuming an even sex ratio, the maximum population growth rate was assumed to be less than 3. The shape parameters (c in (6), and q in (7)) were assumed to be between 0 and 10 to span a wide range of possible dynamics. The coefficient of variation (v) was assumed to be between 0 and 0.4, which was reasonable given experience of modelling other organisms. For the sake of simplicity and to reflect a relative lack of prior information, uniform prior distributions were assumed for all the parameters.

Exact solutions for the probability of transition from one population size to any other population size in the next year can be obtained from the probabilities of the negative binomial distribution. By constructing a matrix of these transition probabilities, and raising the matrix to the power corresponding to the number of years of interest, risks of extinction [Ludwig, 1996b] and the expected minimum population size [McCarthy and Thompson in press] within 100 years were obtained for a small isolated population with an equilibrium population size of 30 females.

A Monte Carlo procedure outlined by Hilborn and Mangel [1997, p. 257] was used to calculate the posterior distributions of the predictions from each of the models (5)-(7). Bayesian confidence intervals on the predictions were calculated for each model separately and also by integrating across the three models.

3.2. Results

For the small isolated population of *Burramys parvus* that we considered, the predicted risks of population extinction within the next 100 years are uncertain, with the 95% confidence intervals for the posterior predictions spanning more than 6 orders of magnitude. Despite such wide confidence intervals, the results in figure 1 demonstrate that the risk of extinction is likely to be less than 0.01.

Figure 2 shows that the expected minimum population size is likely to be between about 8 and 17 females. Moreover, the three different models made similar predictions, with the result that the average was similar to the individual predictions.

3.3. Interpretation

In this example, the confidence intervals for the predictions of the different models were very similar, such that the model-averaged predictions were similar to any one of the predictions for the individual models. This result occurred because these three models happened to make similar predictions, even though they are capable of quite different dynamics. It is conceivable that other models would produce different predictions, such as if an Allee effect or delayed density dependence was included [Burgman et al., 1993]. In this case, the BMA would tend to widen the confidence interval, as was the case for the greater glider habitat models (above).

The low risk of extinction within 100 years (<0.01) for such a small population (30 females as the initial population size) may, at first, seem unlikely. However, it appears that populations of this approximate size have persisted for up to 10,000 years [see McCarthy and Broome, 2000]. Thus, the low risks of extinction appear to be reasonable.

4. CONCLUSIONS

For simplicity, we have highlighted one approach to incorporating model uncertainty in predictions. There would be great value in a detailed study that compared the performance of the various methods available for treating model uncertainty in an ecological context. In particular, a thorough investigation into the value of MCMC, hierarchical and bootstrap approaches would be beneficial.

Whilst we have not demonstrated that BMA provides models that give substantially more accurate predictions, we have demonstrated that BMA provides a simple means for a more satisfactory treatment of prediction uncertainty than currently accepted methods. This may present a case for a compromise between the single-best model and a model averaging approach where

point predictions might be derived from the single-best model and prediction intervals calculated using model averaged variances.

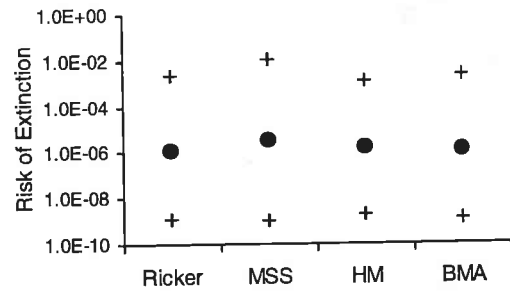


Figure 1. Risk of extinction within the next 100 years for *Burramys parvus* predicted by the four modelling approaches, showing the median (circles) and 95% confidence limits (crosses).

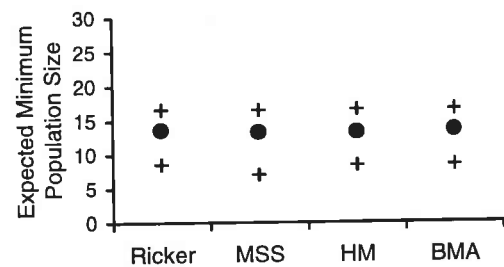


Figure 2. Expected minimum population size within the next 100 years for *Burramys parvus* predicted by the four modelling approaches, showing the median (circles) and 95% confidence limits (crosses).

5. ACKNOWLEDGEMENTS

This work was supported by Queensland's Department of Natural Resources, State Forests of NSW and the Holdsworth Foundation. Jane Elith provided excellent programming assistance and valuable comments.

6. REFERENCES

- Agresti, A., An introduction to categorical data analysis, John Wiley and Sons, New York, 1996.
- Braithwaite, L.W., Studies of the arboreal marsupial fauna of eucalypt forests being harvested for woodpulp at Eden, New South Wales. I. The species and distribution of animals. *Australian Wildlife Research*, 10, 219-229, 1983.
- Buckland, S.T., K.P. Burnham, and N.H. Augustin, Model selection: an integral part of inference. *Biometrics*, 53, 603-618, 1997.

- Burgman, M.A., S. Ferson, and H.R. Akcakaya, Risk assessment in conservation biology, Chapman and Hall, London, 1993.
- Chatfield, C., Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society Series A*, 158, 419-466, 1995.
- Cork, S.J., and P.C. Catling, Modelling distributions of arboreal and ground-dwelling mammals in relation to climate, nutrients, plant chemical defences and vegetation structure in the eucalypt forests of southeastern Australia. *Forest Ecology & Management* 85, 163-175, 1996.
- Cork, S.J., P.C. Catling, L.W. Braithwaite, and D.M. Spratt, Modelling distributions of arboreal and ground living mammals in relation to climate, nutrients, plant chemical defences and vegetation structure in the eucalypt forests of southeastern Australia. in: *International forest biodiversity conference*, Canberra, 1994.
- Davey, S.M., The environmental relationships of arboreal marsupials in a eucalypt forest: a basis for Australian forest wildlife management, PhD Thesis, ANU, 1989.
- Draper, D., Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B-Methodological*, 57, 45-97, 1995.
- ESRI, ArcView 3.1., Environmental Systems Research Institute Inc., Redlands, CA, 1998.
- Hilborn, R. and M. Mangel, The ecological detective: confronting models with data, Princeton University Press, Princeton, New Jersey, 1997.
- Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky, Bayesian model averaging: a tutorial. *Statistical Science* 14, 382-401, 1999.
- Hosmer, D.W., and S. Lemeshow, Applied logistic regression, John Wiley & Sons, New York, 1989.
- Hosmer, D.W., and S. Lemeshow, Confidence interval estimates of an index of quality performance based on logistic regression models. *Statistics in Medicine* 14, 2161-2172, 1995.
- Kavanagh, R.P., Ecology and management of large forest owls in south-eastern Australia., Ph.D, University of Sydney, 1997.
- Kavanagh, R.P., and K.L. Bamkin, Distribution of nocturnal forest birds and mammals in relation to the logging mosaic in south-eastern New South Wales, Australia. *Biological Conservation*, 71, 41-53, 1995.
- Lindenmayer, D.B., R.B. Cunningham, M.T. Tanton, A.P. Smith, and H.A. Nix, Habitat requirements of the mountain brushtail possum and the greater glider in the montane ash-type eucalypt forests of the Central Highlands of Victoria Australia. *Australian Wildlife Research* 17, 467-478, 1990.
- Ludwig, D., Uncertainty and the assessment of extinction probabilities. *Ecological Applications*, 6, 1067-1076, 1996a.
- Ludwig, D., The distribution of population survival times. *American Naturalist*, 147, 506-526, 1996b.
- Madigan, D., and A.E. Raftery, Model selection and accounting for model uncertainty in graphical models using Occams window. *Journal of the American Statistical Association*, 89, 1535-1546, 1994.
- McCarthy, M.A., Red kangaroo (*Macropus rufus*) dynamics: effects of rainfall, harvesting, density dependence and environmental stochasticity. *Journal of Applied Ecology*, 33, 45-53, 1996.
- McCarthy, M.A., and L.S. Broome, A method for validating stochastic models of population viability: a case study of the mountain pygmy-possum (*Burramys parvus*). *Journal of Animal Ecology*, 69, 599-607, 2000.
- McCarthy, M.A., and C. Thompson, Expected minimum population size as a measure of threat. *Animal Conservation*, in press.
- McKay, G.M., Greater Glider, pp. 240-241 in: *Mammals of Australia*, 2nd edition, R. Strahan (ed.), Reed Books, Chatswood, 1995.
- Pausas, J.G., L.W. Braithwaite, and M.P. Austin, Modelling habitat quality for arboreal marsupials in the south coastal forests of New South Wales, Australia, *Forest Ecology & Management*, 78, 39-49, 1995.
- Pearce, J., and S. Ferrier, Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133, 225-245, 2000.
- Raftery, A.E., Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, 83, 251-266, 1996.
- Raftery, A.E., D. Madigan, and C.T. Volinsky, Accounting for model uncertainty in survival analysis improves predictive performance, pp. 323-349 in: *Bayesian Statistics*, J. Bernardo, J. Berger, A. Dawid, and A. Smith (eds), Oxford University Press, 1996.
- Taylor, B.L., The reliability of using population viability analysis for risk classification of species. *Conservation Biology*, 9, 551-558, 1995.