EURING PROCEEDINGS

Comparison of model building and selection strategies

**Paul F. Doherty · Gary C. White ·
Kenneth P. Burnham**

Abstract One challenge an analyst often encounters when dealing with complex mark–recapture models is how to limit the number of a priori models. While all possible combinations of model structures on the different parameters (e.g., ϕ, p) can be considered, such a strategy often results in a burdensome number of models, leading to the use of ad hoc strategies to reduce the number of models constructed. For the Cormack–Jolly–Seber data type, one example of an ad hoc strategy is to hold a general ϕ model structure constant while investigating model structures on p, and then to hold the resulting best structure on p constant and investigate structures on ϕ. Many comparable strategies exist. The effect of following ad hoc strategies on parameter estimates as well as for variable selection and whether model averaging can ameliorate any problems are unknown. By means of a simulation study, we have investigated this informational gap by comparing the all-combinations model building strategy with two ad hoc strategies and with truth, as well as considering the results of model averaging. We found that model selection strategy had little effect on parameter estimator bias and precision and that model averaging did improve bias and precision slightly. In terms of variable selection (i.e., cumulative Akaike's information criterion weights), model sets based on ad hoc strategies did not perform as well as those based on all combinations, as less important variables often had higher weights with the former than with the all possible combinations strategy. Increased sample size resulted in increased variable weights, with an infinite sample size resulting in all variable weights equaling 1 for variables with any predictive influence. Thus, the distinction between statistical importance (dependent on sample size) and biological importance must be recognized when utilizing cumulative weights. We recommend that all-combinations model strategy and model averaging be used. However, if an ad hoc strategy is relied upon to reduce the computational demand, parameter estimates will generally be comparable to the all-combinations strategy, but variable weights will not correspond to the all-combinations strategy.

Keywords Akaike's information criterion · All possible models · Model averaging · Model selection · Multi-model inference · Step-down selection · Stepwise selection

Introduction

Model selection is central to most statistical analyses, especially those based on mark–re-encounter data, and has received much attention (Burnham and Anderson 2002, 2004; Link and Barker 2006; Murray and Conner 2009). However, before model selection can take place, a model set must be available. Model sets can be constructed a priori or in concert with model selection activities, such as in the case of stepwise procedures. Mark–re-encounter problems have the added complexity of having multiple parameters of interest (e.g., apparent survival ϕ, and detection p) in the same likelihood function, but with each parameter modeled as a function of separate independent variables. This complication can be an additional challenge to the analyst because model selection occurs on the entire model (i.e., the likelihood), but biologists often think of

P. F. Doherty (✉) · G. C. White · K. P. Burnham
Department of Fish, Wildlife and Conservation Biology,
Colorado State University, Fort Collins, CO 80523-1474, USA
e-mail: paul.doherty@colostate.edu

each parameter separately. Indeed, the design matrix associated with such models often contains no overlap of lower level parameters (i.e., the independent variables, or columns in the design matrix or β parameters in Program MARK) across higher level parameters (e.g., ϕ and p). However, some exceptions do exist, such as in modeling initial capture–recapture probabilities in closed capture models (White 2008). Few guidelines, especially with solid theoretical foundations, exist to guide the analyst in how to construct and select models in these cases. Ultimately the analyst is interested in which variables are biologically important as well as associated estimates and precision. Here, we explore and test past suggestions on how to approach the interplay between model set construction and model selection in a mark–recapture context.

Model set construction and selection approaches

Burnham and Anderson (2002) advocate that analysts focus on developing a concise model set (4–20 models) that represents scientific hypotheses while recognizing that some practical problems may have closer to 100 models to be considered and that the number of models is often larger with larger data sets; however, the number of models should not exceed the sample size (Anderson 2008). Burnham and Anderson (2002) also suggest that examining "all possible models", especially with a stepwise selection approach, should be avoided due to the high possibility of spurious results.

These suggestions have been heeded by many mark–recapture data analysts. This advice on building candidate model sets and model selection also align with historical approaches (i.e., Lebreton et al. 1992, discussed below) except for the advice against stepwise approaches. However, the term "all possible models" would technically be an infinite set and would also include truth. In practice "all possible models" is usually defined as being conditional on combinations (e.g., additive only, only two-way interactions included) of a limited set of predictor variables. We assume that all variables under consideration potentially have relevance and use the term "all combinations".

With the availability of highly structured mark–recapture models, large data sets, and available computer power, performing a reasonably complete data analysis often conflicts with Burnham and Anderson's (2002) advice. For example, consider a Barker model with higher level parameters S, p, r, R, R', F, and F' (Barker 1997). If an analyst were to consider only two possible hypotheses (or lower level model structures) for each parameter, a possible 128 higher level parameter combinations could be constructed, excluding any interactions. With three structures

on each parameter, a possible 2,187 combinations would result—a much larger model set than Burnham and Anderson (2002) suggest is useful. We believe that biologists will often have more than two or three hypotheses in which they will be interested for each dependent parameter and often have structures including both main and interactive effects; consequently, such high numbers of hypotheses/models are not unusual, especially in non-manipulative experimental studies. With the realized flexibility of multi-state models (Lebreton et al. 2009), the number of higher level parameters in models is essentially unlimited. Many data sets are large enough to easily support hundreds of lower level parameters, and we have been involved with data sets exceeding 187,000 individuals (e.g., Doherty et al. 2002). With available computer power, practical constraints to constructing large numbers of models become fewer each year. Yet little attention has been paid to addressing how model construction and selection should take place in these common, complex problems where a reasonable model set can contain hundreds of models. Should all possible combinations be constructed? Or should ad hoc historical approaches (discussed below) continue to be adopted, although such approaches were developed when complex model structures and powerful computers were not available?

The most often followed historical approach was presented in Lebreton et al.'s (1992) monograph and focused on estimating apparent survival (ϕ) and re-encounter probabilities (p). Lebreton et al. (1992) suggested a "step-down" approach where the model structure on the survival parameter is fixed at a high dimensionality and a model for capture probabilities is selected. "Step-down" is a type of "stepwise" approach. Once a model structure for capture probabilities is settled upon, survival parameters are then focused upon (e.g., Dugger et al. 2006). Lebreton et al. (1992) also suggest that neighboring models should be checked, and many approaches to doing this have been developed, such as keeping highly ranked sub-structures for further comparisons (e.g., Sandercock and Jaramillo 2002). Lebreton et al. (1992) noted that the reverse approach could also be taken (i.e., first keep capture probabilities fixed with a high dimensionality and then focus on modeling the survival parameter), but they slightly favored the former order of operations. They also recognized that these two approaches may not converge to the same model for a particular data set. Lebreton et al. (1992) provided no guidance for the "step-down" order when more than two higher-order parameters occur in a model, but see Sandercock and Jaramillo (2002) and Doherty et al. (2002) for example ad hoc strategies.

Burnham and Anderson (2002) also promote model-averaged parameter estimates where the average estimate

$(\hat{\bar{Y}})$ is calculated by summing the individual model (i) estimates (Y_i) multiplied by the associated Akaike's information criterion (AIC) model weight (w_i) across the model set (R), such that $\hat{\bar{Y}} = \sum_{i=1}^{R} w_i \hat{y}_i$. Lukacs et al. (2010) demonstrated that model averaging over a set of models can protect against spurious results as compared to examining a single top model. In such cases, a balanced model set (i.e., a set where each variable occurs the same number of times) is often desired. However, Lukacs et al. (2010) recommend against using an "all possible models" approach even though this is the strategy they used in demonstrating their main result.

In addition to higher level parameter estimates, analysts are interested in the relative importance of lower level variables. Variables in a top ranked model are often interpreted as being most important. However, Burnham and Anderson (2002) also promote calculating a cumulative AIC weight for each variable by summing the weights from all models in a set that contain a variable of interest.

Given the lack of theoretical development and availability of many ad hoc model set construction/selection strategies, our goal was to explore three different model construction and selection strategies through a simulation approach in a Cormack–Jolly–Seber (CJS) mark–recapture context. We focused on the apparent survival first, capture probability first, and all-combinations approaches, as well as on each approach combined with a model-averaging routine for a total of six strategies. We examined factor selection using bias and precision of parameter estimates [i.e., relative sums of squares (RSS)] as well as variable importance [i.e., cumulative AIC (AIC_c) weights].

Methods

We needed to limit our simulations to a manageable size because the topic of model set construction and selection applies to a wide range of situations. We focused on simulating parameter and covariate values based on a CJS model to estimate ϕ's and p's. We considered a study with 15 occasions (t) and thus 14 ϕ's and 14 p's. We set the mean logit(ϕ) = 1.386 with standard deviation (SD) = 0.5 (approximately $\bar{\phi} = 0.8$ with SD = 0.1), and mean logit (p) = 0.405 with SD = 0.5 (approximately $\bar{p} = 0.6$ with SD = 0.15). Our true generating model was thus (ϕ_t, p_t). We chose this as our generating model to reflect a realized system in which "truth" is complex and the biologist will not have enough covariates to explain all of the variation in the system.

We also assume a situation in which a biologist has identified four variables for each of the parameters as being potentially important. Thus, we considered the ϕ's to be a function of four independent variables ($x_{1\phi}-x_{4\phi}$) and the p's to be a function of four additional independent variables ($x_{1p}-x_{4p}$; Table 1). The mean value for each of the eight covariates was 4 with SD = 1. For truth, we chose to have the four covariates (x_1-x_4) for each variable (ϕ, p) to have a decreasing correlation with the variable (Table 1), but these covariates were not part of the generating model (ϕ_t, p_t) and did not explain all the variance in these parameters. To gain some breadth to our simulation efforts we chose to use three different sets of parameters that fulfilled the above criteria (Table 1).

For each of the three parameter sets, we simulated a release of 100 individuals at occasion 1, and 20 releases on each succeeding occasion to maintain approximately a population of 100 animals. One thousand data sets were constructed from each of the three parameter sets. In addition, we performed limited simulations with the three parameter sets where the number of individuals released was increased by tenfold and 100-fold, respectively. These additional simulations were conducted to confirm the expected behavior of the model selection strategies as sample size increases.

We confronted each of the 1,000 simulated data sets with three different model construction/selection strategies using the model selection criterion AIC_c (Burnham and Anderson 2002). The first model strategy consisted of keeping the model on ϕ at the full additive covariate dimensionality (i.e., $x_{1\phi} + x_{2\phi} + x_{3\phi} + x_{4\phi}$), constructing every possible additive covariate model on p, and then choosing the top p model. We then fixed the structure on p according to the top model and constructed the full complement of covariate models on ϕ, including the constant (i.e., "dot") model with no covariates. This strategy is labeled "p first" and resulted in a model set containing 31 models. In the second strategy, we kept the model on p at the full additive covariate dimensionality and constructed every possible additive covariate model on ϕ and chose the top ϕ model. We then fixed the structure on ϕ according to the top model and constructed the full complement of covariate models on p, including the constant (i.e., "dot") model with no covariates. This strategy is labeled "ϕ first" and resulted in a model set containing 31 models. The third strategy was to construct every possible covariate combination of ϕ and p structure with each other, for a total of 256 models (labeled "all combinations"). These three strategies each produced a minimum AIC_c model as well as a model set from which we produced model averaged estimates of ϕ and p. We focus on the minimum AIC_c model estimates as well as model-averaged estimates in the Results. For comparative purposes we also constructed the (ϕ_t, p_t) model for each data set, but this model was not considered in the model set.

Table 1 Realized correlations between the eight covariates and the logit and back-transformed parameter values of ϕ and p for the three parameter sets

	ϕ	Logit(ϕ)	$x_{1\phi}$	$x_{2\phi}$	$x_{3\phi}$		p	Logit(p)	x_{1p}	x_{2p}	x_{3p}
Parameter set 1											
Logit(ϕ)	0.983	1.000				Logit(p)	0.997	1.000			
$x_{1\phi}$	0.834	0.861	1.000			x_{1p}	0.911	0.917	1.000		
$x_{2\phi}$	0.595	0.645	0.528	1.000		x_{2p}	0.666	0.660	0.456	1.000	
$x_{3\phi}$	0.239	0.266	0.037	−0.130	1.000	x_{3p}	0.302	0.325	0.325	0.043	1.000
$x_{4\phi}$	−0.055	0.052	0.013	0.274	−0.311	x_{4p}	−0.060	−0.013	−0.067	0.017	0.385
Parameter set 2											
Logit(ϕ)	0.958	1.000				Logit(p)	0.997	1.000			
$x_{1\phi}$	0.893	0.864	1.000			x_{1p}	0.906	0.900	1.000		
$x_{2\phi}$	0.313	0.484	0.121	1.000		x_{2p}	0.462	0.491	0.143	1.000	
$x_{3\phi}$	0.306	0.262	0.042	0.018	1.000	x_{3p}	0.198	0.228	0.276	−0.115	1.000
$x_{4\phi}$	0.081	0.058	−0.136	0.023	0.378	x_{4p}	0.003	0.052	0.031	0.262	0.030
Parameter set 3											
Logit(ϕ)	0.977	1.000				Logit(p)	0.998	1.000			
$x_{1\phi}$	0.900	0.865	1.000			x_{1p}	0.818	0.817	1.000		
$x_{2\phi}$	0.575	0.706	0.387	1.000		x_{2p}	0.459	0.470	−0.051	1.000	
$x_{3\phi}$	0.002	−0.005	−0.130	−0.091	1.000	x_{3p}	0.247	0.225	0.038	0.266	1.000
$x_{4\phi}$	0.033	0.063	−0.170	0.295	−0.223	x_{4p}	0.058	0.076	0.103	−0.106	−0.563

ϕ Apparent survival, p re-encounter probabilities, $x_{1\phi}$–$x_{4\phi}$ four independent variables of which ϕ is a function, x_{1p}–x_{4p} four independent variables of which p is a function

To evaluate how each model selection strategy performed, for each data set and model selection strategy we computed the RSS for ϕ and p parameter estimates in the minimum AIC_c model as well as for the model-averaged estimates as:

$$\text{RSS}(\theta) = \sum_{i=1}^{13} \left(\frac{\hat{\theta}_i - \theta}{\theta} \right)^2$$

The closer the RSS value is to zero, the less bias and more precise are the estimates. We then calculated averages and standard deviations across the 1,000 repetitions (Table 2) and also summed the RSS values associated with the ϕ and with the p structures to obtain an overall RSS for the model (Table 2). For comparative purposes, in Table 2 we also identify the minimum RSS model from each data set and present the RSS values from the (ϕ_t, p_t) model. The RSS values were computed only for the first 13 estimates of ϕ and p because we wanted to be able to compare RSS values from our various model selection strategies against the fully time-specific model (ϕ_t, p_t) in which the last ϕ and p estimates are confounded and hence not usefully estimated.

In addition to bias and the precision of estimates, we were also interested in overall variable selection. We thus calculated the cumulative AIC_c weight for the 8 variables in each model set to evaluate variable importance (Burnham and Anderson 2002) and averaged these cumulative weight values across the 1,000 repetitions.

Results

Estimator bias and precision

The three parameter sets produced similar RSS results and rankings of the model selection strategies. We found that the three model selection strategies performed nearly identically in selecting minimum AIC_c models with similar RSS values (Table 2). Model-averaged estimates always had smaller RSS values (2–6%) and also had a smaller SD, which includes model selection uncertainty (Table 2). For all three sets of parameters simulated, none of the minimum AIC_c models selected by any of the three strategies approached the minimum RSS values possible, but all three model selection strategies were approximately the same distance from this theoretically best RSS value. Thus, we would expect that any of these three strategies would produce almost equivalent predictions of the parameter values but that model-averaging would be advantageous. The models selected by AIC_c also had fewer parameters than the minimum RSS model. With larger sample sizes we found model selection results closer to the minimum RSS model.

RSS values for the fully time-specific models were larger than any of the model selection strategies for the sample sizes used here because of the poor precision obtained from estimating 27 parameters. However, as sample size increased, the fully time-specific model produced RSS values smaller than any of those obtained from

Table 2 Model selection simulation results for each of the three sets of parameters

Strategy[a]	Criterion[b]	Average RSS				SD RSS			
		ϕ	p	$\phi + p$	# Par.[c]	ϕ	p	$\phi + p$	# Par.
Parameter set 1									
	Min RSS	0.078	0.202	0.280	8.0	0.035	0.061	0.068	1.0
	ϕ_t, p_t	0.186	0.337	0.523	27.0	0.093	0.172	0.210	0.0
p first	Min AIC$_c$	0.109	0.254	0.363	6.6	0.062	0.097	0.116	1.0
	ModelAve	0.102	0.251	0.353	–	0.054	0.092	0.107	–
ϕ first	Min AIC$_c$	0.110	0.257	0.367	6.6	0.063	0.099	0.119	1.0
	ModelAve	0.106	0.256	0.362	–	0.059	0.091	0.110	–
All combinations	Min AIC$_c$	0.109	0.256	0.366	6.6	0.062	0.098	0.118	1.0
	ModelAve	0.103	0.252	0.355	–	0.055	0.087	0.102	–
Parameter set 2									
	Min RSS	0.051	0.121	0.172	7.9	0.016	0.044	0.046	1.1
	ϕ_t, p_t	0.101	0.195	0.296	27.0	0.053	0.104	0.124	0.0
p first	Min AIC$_c$	0.073	0.147	0.220	7.0	0.028	0.064	0.069	1.1
	ModelAve	0.065	0.145	0.210	–	0.027	0.061	0.066	–
ϕ first	Min AIC$_c$	0.072	0.147	0.219	6.9	0.028	0.064	0.069	1.1
	ModelAve	0.069	0.144	0.214	–	0.027	0.058	0.064	–
All combinations	Min AIC$_c$	0.073	0.147	0.220	6.9	0.028	0.064	0.069	1.1
	ModelAve	0.065	0.143	0.208	–	0.026	0.057	0.062	–
Parameter set 3									
	Min RSS	0.026	0.067	0.093	7.7	0.012	0.025	0.028	1.0
	ϕ_t, p_t	0.089	0.153	0.200	27.0	0.042	0.067	0.086	0.0
p first	Min AIC$_c$	0.039	0.084	0.123	6.6	0.022	0.038	0.045	1.1
	ModelAve	0.034	0.082	0.115	–	0.021	0.037	0.044	–
ϕ first	Min AIC$_c$	0.039	0.084	0.123	6.7	0.022	0.037	0.044	1.1
	ModelAve	0.036	0.081	0.118	–	0.021	0.036	0.042	–
All combinations	Min AIC$_c$	0.039	0.084	0.123	6.6	0.022	0.037	0.044	1.1
	ModelAve	0.034	0.080	0.114	–	0.020	0.036	0.043	–

RSS Relative sum of squares of the parameter estimates, SD standard deviation

[a] Each parameter set was used to produce 1,000 data sets. The p-first, ϕ-first, and all-combinations modeling strategies were applied to each data set (see text for description)

[b] The model with the minimum RSS (Min RSS) and (ϕ_t, p_t) model are also presented for comparative purposes. RSS was calculated for the minimum cumulative Akaike's information criterion model (Min AIC$_c$) as well as for the model-averaged estimates (ModelAve)

[c] # Par., Average number of parameters across the 1,000 repetitions

models based on the eight covariates. This result is because none of the eight covariates has a perfect correlation with true parameter values, whereas the data were generated from the fully time-specific model.

Variable selection

In contrast, the three model selection strategies did not perform equivalently for determining cumulative variable weights (Table 3). In general, covariates highly correlated (e.g., x_1, x_2) with the true parameter values produced similar weights for each of the three model selection strategies. However, for covariates with low correlations (e.g., x_3, x_4) to the true parameter values used to generate

the observed data for the 1,000 simulations, the three model selection strategies produced quite different variable weights. The ϕ-first strategy often produced high weights (>0.5) for the weakly correlated covariates of p (i.e., x_{3p}, x_{4p}), whereas the all-combinations strategy had lower weights and the p-first strategy had the lowest weights for these 2 covariates. The reverse pattern was detected for the low correlated covariate values of ϕ ($x_{3\phi}$, $x_{4\phi}$): the p-first strategy had the highest weights and the ϕ-first strategy had the lowest.

Sample sizes markedly affected these cumulative weights. As sample size increased, the weights for all covariates approached 1, even for covariates with little correlation to the true parameter values.

Table 3 Summary of cumulative AIC$_c$ variable weights for each of the three sets of parameters

Variable	Correlation	p first		ϕ first		All combinations	
		Weight	SD	Weight	SD	Weight	SD
Parameter set 1							
$x_{1\phi}$	0.861	0.832	0.198	0.825	0.306	0.814	0.212
$x_{2\phi}$	0.645	0.714	0.224	0.652	0.391	0.677	0.241
$x_{3\phi}$	0.266	0.618	0.223	0.551	0.395	0.568	0.228
$x_{4\phi}$	0.052	0.493	0.159	0.306	0.329	0.429	0.158
x_{1p}	0.917	1.000	0.000	1.000	0.000	1.000	0.000
x_{2p}	0.66	0.986	0.085	0.973	0.076	0.962	0.096
x_{3p}	0.325	0.215	0.323	0.524	0.151	0.380	0.141
x_{4p}	−0.013	0.21	0.319	0.52	0.152	0.380	0.144
Parameter set 2							
$x_{1\phi}$	0.864	1.000	0.001	1.000	0.000	1.000	0.001
$x_{2\phi}$	0.484	0.741	0.224	0.724	0.367	0.693	0.241
$x_{3\phi}$	0.262	0.691	0.223	0.637	0.398	0.635	0.241
$x_{4\phi}$	0.058	0.538	0.177	0.299	0.352	0.427	0.174
x_{1p}	0.9	1.000	0.000	1.000	0.000	1.000	0.000
x_{2p}	0.491	0.904	0.212	0.891	0.161	0.878	0.174
x_{3p}	0.228	0.243	0.295	0.49	0.154	0.383	0.147
x_{4p}	0.052	0.292	0.334	0.507	0.173	0.403	0.169
Parameter set 3							
$x_{1\phi}$	0.865	0.987	0.052	0.993	0.057	0.986	0.057
$x_{2\phi}$	0.706	0.708	0.235	0.679	0.373	0.678	0.247
$x_{3\phi}$	−0.005	0.461	0.156	0.253	0.303	0.382	0.149
$x_{4\phi}$	0.063	0.538	0.21	0.413	0.384	0.478	0.216
x_{1p}	0.817	0.999	0.023	1.000	0.011	0.999	0.013
x_{2p}	0.47	0.977	0.11	0.962	0.088	0.951	0.107
x_{3p}	0.225	0.281	0.365	0.546	0.176	0.421	0.177
x_{4p}	0.076	0.246	0.341	0.529	0.163	0.400	0.161

Each parameter set was simulated for 1,000 realizations. Correlations are between the covariate and the true parameter value used to generate the observed data for each of the 1,000 simulations upon which this table is based. Weights are the average cumulative AIC$_c$ weights for each of the eight covariates computed from the 1,000 data sets for each of the three model selection strategies

Discussion

The three modeling strategies evaluated here performed almost equivalently in terms of RSS, but model averaging improved all three strategies. Thus, we recommend that model averaging be routinely considered. The three strategies did not perform equivalently in terms of variable weights, and a balanced, all-combinations strategy is probably best. The p-first and ϕ-first strategies placed more cumulative weight on lowly correlated covariates associated with ϕ and p, respectively. Part of the reason for this incompatibility is that each variable was not included in the same number of models in the model set (i.e., this was an unbalanced model set) for the ϕ-first and p-first strategies and the parameter initially kept at a high dimensionality was "forced" to explain variation in the data. Unfortunately, the cumulative weights on the lowly correlated variables from the p-first and ϕ-first strategies often exceeded 0.5. Barbieri and Berger (2004) suggest that variables with cumulative weight >0.5 are the most important and should be focused upon. If this advice is followed, spurious results could occur under the p-first and ϕ-first strategies. Other ad hoc strategies could be considered. Keeping p or ϕ constant, rather than at a high dimensionality, is a possibility, but this strategy will perform poorly because the variation explained will be forced onto the other parameter and will lead to spurious variable weighting results. We focused on CJS-type data. More complicated models (e.g., Barker model, multi-state, robust design) with more parameters present additional difficulties for ad hoc strategies as deciding on the order of parameters to model becomes tenuous. By following the all-combinations strategy, such a problem does not exist. We suggest that the all-combinations selection strategy, which is

balanced for each variable, seems to be the method that is best for achieving useful estimates of variable weights. However, untested strategies to address unbalanced model sets do exist. One suggestion (D. MacKenzie personal communication) is to use $[w/(1-w)]/[f/(1-f)]$ as a measure of support where w is the cumulative AIC weight for a variable and f is the frequency of the variable in the model set. Values $\gg 1$ would indicate support for importance, values near 1 would be inconclusive, and values $\ll 1$ would indicate little support for importance. In this case, the range of possible values would be 0 to infinity.

The effect of increasing sample size on variable weights is to drive all variable weights to 1. That is, a very large (or asymptotic) sample size will result in all the weight being on the model that contains all of the predictors. In fact, Murray and Conner (2009) also demonstrate this behavior by simulating data sets with large numbers of individuals (i.e., 20,000) such that the most parameterized model contained all the weight. In this case, cumulative variable weights are of little use in terms of ranking variable importance because all variables appear to be equally important. A similar problem occurs with the hypothesis testing paradigm, in that a test is statistically significant ($P < 0.001$), yet the effect size for the variable is biologically irrelevant. Neither paradigm allows the user to select only biologically relevant variables as both paradigms are evaluating the statistical relevance of the variables. In practice, users seldom have such large sample sizes that this asymptotic behavior creates a problem, although the sample size can be misconstrued in mark–recapture problems. The sample size in such problems is closer to the number of releases, and not the number of individuals.

We selectively chose the three parameter sets to provide a range of correlations, conditional on our criteria, to be able to assess performance of the three model selection strategies. In theory, one could generate data sets in which a true correlation matrix was specified, and not limit the parameter space to just three sets of parameters. However, in practice, realized correlations between the covariates and true parameter values are often quite different than the expected correlations given the small sample of observations used in these CJS models. We screened >1,000 sets of parameters to obtain these three parameter set that provided a good range of realized correlations. Additional theoretical developments and simulation work focusing on different values of p and ϕ, different correlation structures, and more complicated mark–recapture models (e.g., multi-state models) would be useful, as we view our work as a small step towards developing useful model set construction/selection guidelines for practitioners.

References

Anderson DR (2008) Model based inference in the life sciences—a primer on evidence. Springer, New York

Barbieri MM, Berger JO (2004) Optimal predictive model selection. Ann Stat 32:870–897

Barker RJ (1997) Joint modeling of live-recapture, tag-resight, and tag-recovery data. Biometrics 53:666–677

Burnham KP, Anderson DR (2002) Model selection and multimodel inference—a practical information-theoretic approach, 2nd edn. Springer, New York

Burnham KP, Anderson DR (2004) Multimodel inference—understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304

Doherty PF, Nichols JD, Tautin J, Voelzer JF, Smith GW, Benning DS, Bentley VR, Bidwell JK, Bollinger KS, Brazda AR, Buelna EK, Goldsberry JR, King RJ, Roetker FH, Solberg JW, Thorpe PP, Wortham JS (2002) Sources of variation in breeding-ground fidelity of mallards (*Anas platyrhynchos*). Behav Ecol 13:543–550

Dugger KM, Ballard G, Ainley DG, Barton KJ (2006) Effects of flipper bands on foraging behavior and survival of Adelie Penguins (*Pygoscelis adeliae*). Auk 123:858–869

Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals—a unified approach with case-studies. Ecol Monogr 62:67–118

Lebreton JD, Nichols JD, Barker RJ, Pradel R, Spendelow JA (2009) Modeling individual animal histories with multistate capture-recapture models. Adv Ecol Res 41:88–173

Link WA, Barker RJ (2006) Model weights and the foundations of multimodel inference. Ecology 87:2626–2635

Lukacs PM, Burnham KP, Anderson DR (2010) Model selection bias and Freedman's paradox. Ann Inst Stat Math 62:117–125

Murray K, Conner MM (2009) Methods to quantify variable importance: implications for the analysis of noisy ecological data. Ecology 90:348–355

Sandercock BK, Jaramillo A (2002) Annual survival rates of wintering sparrows: assessing demographic consequences of migration. Auk 119:149–165

White GC (2008) Closed population estimation models and their extensions in program MARK. Environ Ecol Stat 15:89–99