

# Plausible combinations: An improved method to evaluate the covariate structure of Cormack-Jolly-Seber mark-recapture models

Jeffrey F. Bromaghin<sup>1</sup>, Trent L. McDonald<sup>2</sup>, Steven C. Amstrup<sup>3</sup>

<sup>1</sup>US Geological Survey, Alaska Science Center, Anchorage, USA; [jbromaghin@usgs.gov](mailto:jbromaghin@usgs.gov)

<sup>2</sup>Western Ecosystems Technology, Inc., Cheyenne, USA; [tmcdonald@west-inc.com](mailto:tmcdonald@west-inc.com)

<sup>3</sup>Polar Bears International, Bozeman, USA; [samstrup@pbears.org](mailto:samstrup@pbears.org)

Received 18 November 2012; revised 20 December 2012; accepted 30 December 2012

## ABSTRACT

Mark-recapture models are extensively used in quantitative population ecology, providing estimates of population vital rates, such as survival, that are difficult to obtain using other methods. Vital rates are commonly modeled as functions of explanatory covariates, adding considerable flexibility to mark-recapture models, but also increasing the subjectivity and complexity of the modeling process. Consequently, model selection and the evaluation of covariate structure remain critical aspects of mark-recapture modeling. The difficulties involved in model selection are compounded in Cormack-Jolly-Seber models because they are composed of separate sub-models for survival and recapture probabilities, which are conceptualized independently even though their parameters are not statistically independent. The construction of models as combinations of sub-models, together with multiple potential covariates, can lead to a large model set. Although desirable, estimation of the parameters of all models may not be feasible. Strategies to search a model space and base inference on a subset of all models exist and enjoy widespread use. However, even though the methods used to search a model space can be expected to influence parameter estimation, the assessment of covariate importance, and therefore the ecological interpretation of the modeling results, the performance of these strategies has received limited investigation. We present a new strategy for searching the space of a candidate set of Cormack-Jolly-Seber models and explore its performance relative to existing strategies using computer simulation. The new strategy provides an improved assessment of the importance of covariates and covariate

combinations used to model survival and recapture probabilities, while requiring only a modest increase in the number of models on which inference is based in comparison to existing techniques.

**Keywords:** Capture-Recapture; Survival; Model Building; Model Selection; Model Averaging; Multi-Model Inference; Covariates; Covariate Weights; CJS; Akaike's Information Criterion; AIC; AIC<sub>c</sub>

## 1. INTRODUCTION

Mark-recapture models are among the most widely utilized classes of models in quantitative population ecology. Historically, mark-recapture models were primarily used to estimate abundance [1,2] and survival rates [3,4]. The diversity and complexity of these models have advanced rapidly in recent years, and modern mark-recapture models are now commonly utilized to estimate a variety of vital rates and parameters related to processes such as immigration, state transition, and geographic transience [5-11]. A key factor in the utility of mark-recapture models is the ability to model natural parameters, such as survival rates, as functions of explanatory covariates [12]. The use of covariates substantially enhances the value of mark-recapture models for ecological investigation by providing a mechanism for both generating and testing hypotheses regarding linkages between population vital rates and characteristics of individual animals and their environments.

An unavoidable consequence of using covariates is an increase in the complexity and subjectivity of the modeling process itself. Most mark-recapture experiments are not explicitly designed to test specific hypotheses, but rather generate observational data that may be used to investigate relationships and formulate hypotheses through

exploratory modeling [13]. Mark-recapture investigations are therefore subject to the difficulty of establishing causation with an analysis of observational data [14]. Furthermore, the number of covariates available for inclusion in a model may be large, and ecological data often possess high variance and collinearity. Such properties, in combination with the complexity of ecological processes underlying population vital rates, often complicate the modeling process.

Several aspects of mark-recapture modeling continue to generate considerable interest in the ecological and biometrics literature, including the formation of a candidate set of models. Many investigators advocate developing a relatively small number of models that embody explicit biological or ecological hypotheses [15-17]. Other investigators have noted potential drawbacks with that approach in certain situations. In particular, inference is conditional on the candidate model set, which can cause significant bias if knowledge about the system under study is limited and well-fitting models are not included in the candidate model set [13]. Model formulation is subject to the experience and subjectivity of the investigator, and an overly restrictive model set can also lead to biased estimates [18]. Similarly, ecological processes can be dynamic and an overreliance on prior information to inform model development may mask important ecological relationships. A rigid focus on testing *a priori* hypotheses by definition precludes exploratory modeling that may generate new insights which merit further investigation [13,19]. These varied perspectives are largely encapsulated by model classifications as “white box (theory-based), black box (empirical) and grey box (theory-influenced empirical)” [20], or predictive versus explanatory [21].

The development of a candidate model set is a crucial component of the modeling process that is naturally subject, in part, to the particular objectives of individual investigations. If the primary objective is to test *a priori* hypotheses, developing a relatively small number of models that embody the hypotheses of interest is obviously sound [15-17]. In such investigations, identifying the covariates that seem most strongly associated with natural parameters, *i.e.*, the survival and recapture probabilities of interest rather than covariate coefficients that are the parameters directly estimated, may be of more interest than estimates of those probabilities. In other investigations, the primary objective may be to estimate natural parameters, perhaps to describe and ultimately forecast population dynamics [22], and covariates and their coefficients may or may not be of great interest. Investigations that are more exploratory in nature are likely to incorporate a larger number of covariates into the candidate model set than investigations designed to test a small number of hypotheses.

Once a candidate model set is formed, one must implement a strategy to evaluate the models. The most natural and logically unbiased strategy is to evaluate all candidate models. However, even when candidate models are developed with careful consideration of their biological relevance and plausibility [15-17], the number of models in the set can exceed reasonable limits. Large model sets can easily result when models are formed from combinations of sub-models for different classes of natural parameters, such as survival and recapture probabilities. In addition, the existence of animal classes with potentially unique vital rates or the need to model complex patterns of capture effort through time can substantially increase the number of models to be considered. As an example, a mark-recapture investigation of polar bears (*Ursus maritimus*) in which we were recently involved (unpublished), modeled survival probabilities for four age classes with 24, 10, 1, and 6 sub-models, respectively, totaling 1440 ( $= 24 \times 10 \times 1 \times 6$ ) sub-models for survival. The number of sub-models considered for recapture probabilities was a more modest 96, constructed to incorporate age-class heterogeneity and a complex pattern of capture effort through time. These sub-models were selected from a much larger collection of potential models identified using prior knowledge, expert opinion, and ecological relevance. The model set of all combinations of survival and recapture probability sub-models was therefore 138,240 ( $= 1440 \times 96$ ). Given such a large model space, which is common in mark-recapture investigations, an effective method of searching the model space without estimating the parameters of every candidate model is obviously advantageous.

Several investigators have developed strategies for evaluating a subset of all mark-recapture models within a Bayesian framework. For example, McCrea *et al.* [23] start with a null model and incrementally increase the complexity of component sub-models until a stopping rule is reached, leading to a single preferred model. Note that inference based on a single model increases the risk of model selection bias and precludes the benefits of multi-model inference [15]. King and Brooks [24] use reversible jump Markov chain Monte Carlo to explore a subset of a model space and simultaneously estimate model parameters, which facilitates multi-model inference. Similarly, Sisson and Fan [18] use a simulated annealing algorithm to search a model space and base inference on a subset of high scoring models.

The development of strategies to efficiently search a model space appears to have received less attention outside of a Bayesian framework, and the evaluation of model structure and covariate importance for the Cormack-Jolly-Seber (CJS) mark-recapture model [25,26] remains an important area of inquiry. The CJS model is among the earliest open population models and one of the most

simple structurally, being comprised of sub-models for survival ( $\varphi$ ) and recapture ( $p$ ) probabilities, but nevertheless continues to enjoy widespread application [27-31]. An underappreciated consequence of the statistical dependence between sub-model parameters is that the covariate structure in one sub-model can influence the apparent predictive ability of covariates in the other sub-model, termed information “leak” by Catchpole *et al.* [32], which complicates model evaluation. In addition, it is not unusual for several biologically plausible sub-models for  $p$  and  $\varphi$  to have substantially different covariate structure and yet be similarly supported by data. Consequently, methods used to search a CJS model space can be expected to influence parameter estimation, the assessment of covariate importance, and therefore the ecological interpretation of the modeling results.

The most common strategies to search a CJS model space are the so-called  $p$ - and  $\varphi$ -first strategies [25,26]. In the  $p$ -first strategy, one first searches the  $p$  dimension of the model space by pairing all candidate sub-models for  $p$  with a general and flexible sub-model for  $\varphi$ . One then searches the  $\varphi$  dimension after conditioning on a selected sub-model for  $p$  and pairing it with the remaining sub-models for  $\varphi$ . This stepwise method can be repeated as often as desired, but in practice one or two rounds of pairing a fixed sub-model with a number of candidates for the other sub-model are common. From among the resulting subset of models, a single model may emerge as overwhelmingly dominant and be selected as the preferred model [33] or, more typically, several models will have meaningful support and multi-model inference will be employed [15]. In the  $\varphi$ -first strategy, the order of sub-model evaluation is simply reversed. Variations of these basic strategies are common [3,27,28].

Doherty *et al.* [34] conducted a simulation study to compare the performance of the  $p$ - and  $\varphi$ -first strategies with that of all combinations (AC) of candidate  $p$  and  $\varphi$  sub-models for CJS models. They found that all three strategies performed similarly with respect to estimation of natural parameters, but that the  $p$ - and  $\varphi$ -first strategies provided inferior assessments of covariate importance. If assessing the utility of covariates for modeling  $p$  and  $\varphi$  is a research objective, the AC strategy is preferred to such stepwise procedures [34]. However, as previously noted, the size of the candidate model set in some investigations may render the AC strategy infeasible.

We introduce a new strategy, called the plausible combinations (PC) strategy, and investigate its performance by amending the simulations of Doherty *et al.* [34] to include the PC strategy as an additional alternative. The results reveal that the PC strategy closely mimics the performance of the AC strategy and provides an unbiased assessment of covariate structure, with only a modest increase in the average number of models evaluated

compared to the  $p$ - and  $\varphi$ -first strategies. We recommend this new strategy for situations in which the number of component sub-model combinations is sufficiently large as to render the AC strategy unattractive.

## 2. METHODS

The strategy for model selection and covariate evaluation we propose is a generalization of the  $p$ - and  $\varphi$ -first strategies which provides a more effective search of a model space. The most general sub-model for  $\varphi$  is first paired with all candidate sub-models for  $p$ . The resulting models are then used to identify a list of “plausible” sub-models for  $p$  using a two-step procedure. In the first step, models with an  $AIC_C$  weight [15] exceeding a minimum threshold of, say, 0.01 are identified. These models, termed “high weight” models, provide a reasonable fit to the data and are parsimonious. Models with lower weight may provide an equal or better approximation of the data, but rank low by  $AIC_C$  due to the penalty caused by the number of parameters in the general  $\varphi$  sub-model. Consequently, models with a likelihood exceeding the minimum value of the likelihood among high weight models,  $L_m$ , are identified as “high likelihood” models in the second step. These models approximate the data as well as the high weight models and could become competitive from an  $AIC_C$  perspective if some of the sub-models for  $\varphi$  with fewer parameters provide a good approximation to the data. The sub-models for  $p$  associated with both high weight and high likelihood models are deemed “plausible” for  $p$ . Note that the only product of this two-step procedure is a list of plausible sub-models for  $p$ ; the estimates themselves are discarded and not utilized further. A list of sub-models deemed plausible for  $\varphi$  is similarly developed by reversing the order of  $p$  and  $\varphi$  sub-model conditioning. The parameters of all combinations of plausible sub-models for  $p$  and  $\varphi$  are estimated and inference is based on the resulting subset of the entire model space, an approach termed the plausible combinations (PC) strategy.

The performance of the PC strategy was evaluated by replicating the essential elements of the simulations of Doherty *et al.* [34]. Independent simulations were based on three sets of parameters, each consisting of survival probabilities ( $\varphi$ ), four survival probability covariates ( $x_{1\varphi}$ ,  $x_{2\varphi}$ ,  $x_{3\varphi}$ ,  $x_{4\varphi}$ ), recapture probabilities ( $p$ ), and four recapture probability covariates ( $x_{1p}$ ,  $x_{2p}$ ,  $x_{3p}$ ,  $x_{4p}$ ) for 15 capture occasions (**Table 1**); the parameter sets were kindly provided by Dr. Paul Doherty and Dr. Gary White. A total of 16 sub-models, consisting of a temporally-invariant model containing no covariates (the “.” model) and all 15 possible additive sub-models containing 1, 2, 3, and 4 covariates were evaluated for both  $\varphi$  and  $p$ .

For each parameter set, the covariates were treated as fixed constants and  $\varphi$  and  $p$  (**Table 1**) were used to

**Table 1.** Three parameter sets, each consisting of survival ( $\varphi$ ) and recapture ( $p$ ) probabilities and their four respective covariates for 15 capture occasions.

Occasion	$\varphi$ covariates					$p$ covariates				
	$\varphi$	$X_{1\varphi}$	$X_{2\varphi}$	$X_{3\varphi}$	$X_{4\varphi}$	$p$	$X_{1p}$	$X_{2p}$	$X_{3p}$	$X_{4p}$
Parameter set 1										
1	0.78054	4.10682	3.36879	3.19860	3.77595	-	-	-	-	-
2	0.54653	3.00457	3.91050	2.33771	4.79420	0.36174	3.04087	3.57580	4.30982	6.54941
3	0.69731	3.45072	3.48765	3.53167	4.96847	0.73155	4.76051	4.06977	4.61496	4.82786
4	0.90672	5.14742	4.51119	5.23463	4.78903	0.32954	2.19080	2.98441	4.53637	4.74755
5	0.85285	4.42197	6.12590	1.90597	4.12622	0.43003	4.01534	3.12950	3.45100	4.05404
6	0.85305	3.90633	4.83529	2.52349	5.21135	0.36717	2.94002	3.01845	3.77055	4.70762
7	0.80897	4.62007	3.44390	4.04234	1.83646	0.47062	2.81586	3.83578	4.16405	4.70793
8	0.91751	5.19080	5.48953	4.31543	5.05406	0.63429	4.39905	4.79794	4.22775	4.02451
9	0.77080	4.57286	3.61606	1.75918	3.98934	0.64571	4.73330	3.66544	2.51173	2.70288
10	0.79407	3.50609	4.24305	4.94334	2.54711	0.85093	5.93548	4.29249	5.85046	4.99177
11	0.87805	4.48614	4.88278	5.16065	3.29780	0.65336	3.55556	5.68795	3.48044	3.58899
12	0.80695	3.89087	4.71763	3.94112	3.63841	0.36164	3.16790	3.34932	3.10263	4.22011
13	0.65418	3.15993	3.84917	4.70577	4.09989	0.18730	1.99428	1.91211	4.05147	3.24748
14	0.70558	2.83526	2.47648	4.65518	3.67438	0.60062	4.35061	2.48810	5.26609	4.24447
15	-	-	-	-	-	0.47631	3.06997	3.08562	1.87541	4.60062
Parameter set 2										
1	0.80715	4.15101	3.54239	3.91634	4.76327	-	-	-	-	-
2	0.96775	6.40023	5.78733	3.39285	3.51462	0.73822	5.32159	4.09081	3.23499	4.70665
3	0.55040	2.01089	5.75926	3.46112	3.27696	0.62410	4.03667	3.55712	4.28458	3.78543
4	0.69337	3.94577	3.59877	2.56060	4.17914	0.72735	4.06055	5.05990	4.75615	3.38056
5	0.69579	2.36536	4.04049	4.02345	5.11190	0.44829	2.59276	4.56668	3.91611	4.84179
6	0.87801	5.28450	3.66794	3.65808	2.96759	0.45604	3.67669	3.51874	3.68676	4.02163
7	0.91283	4.71649	5.57875	4.72345	4.09710	0.25670	1.43081	3.91371	4.90297	5.03117
8	0.72801	4.22758	2.48859	4.11427	5.48212	0.85316	5.39869	5.94874	5.49458	5.50698
9	0.85654	4.92854	3.70778	4.32183	3.54931	0.62645	3.01264	5.52889	2.61170	3.04019
10	0.84578	4.81498	4.83275	1.90901	4.10746	0.80853	5.63651	4.96053	5.09375	5.15802
11	0.87960	4.88099	2.73035	5.38649	4.16983	0.67767	4.97605	3.76371	5.74120	1.74198
12	0.92863	4.65170	6.78657	5.53511	5.45633	0.60850	4.22143	3.75905	3.21514	2.37209
13	0.74538	4.23777	4.13487	2.16674	3.35181	0.47810	2.75882	4.59186	3.07973	5.23670
14	0.95844	5.45418	6.19805	3.32524	4.93703	0.47621	3.26942	2.31183	5.26073	3.61626
15	-	-	-	-	-	0.72368	5.60889	3.70113	4.40537	5.51616
Parameter set 3										
1	0.81688	4.64674	3.05472	3.31596	3.83961	-	-	-	-	-
2	0.92647	5.03590	6.05558	3.57843	5.34556	0.68090	4.69925	4.88705	3.51999	4.05543
3	0.82242	4.67945	3.66638	4.62771	2.42649	0.68934	5.45352	1.91894	2.93835	4.89466
4	0.88447	4.42961	4.48874	3.95199	4.97414	0.71687	5.34087	3.23969	5.66851	2.33607
5	0.79864	4.08107	3.99195	4.07886	3.48046	0.44347	3.83043	2.08761	4.18036	4.48236
6	0.76479	3.30962	4.22027	4.83650	4.31012	0.39863	3.22502	3.30458	3.46726	2.51518
7	0.93929	5.24350	5.62082	4.12056	4.48738	0.54487	3.44159	3.94949	3.17209	4.38880
8	0.63162	2.51629	3.86661	4.08282	2.45180	0.51466	3.65704	3.90787	3.93766	4.04720
9	0.68473	2.15965	3.80472	5.02869	4.84598	0.71013	4.69741	4.37175	5.26810	4.78659
10	0.86962	5.23654	3.87029	3.84034	2.46841	0.51633	3.68482	3.13815	3.54855	3.43981
11	0.82458	3.81315	3.32939	4.80756	4.18999	0.80994	5.20251	5.61024	3.16975	5.28320
12	0.93074	4.69994	5.55690	4.62027	2.68545	0.47346	3.97540	2.83237	2.76890	5.56294
13	0.84399	3.65195	4.43899	3.14567	4.06658	0.70629	5.04407	4.37958	4.80395	1.72245
14	0.66032	2.79740	4.19595	2.84050	5.00316	0.64932	3.63556	4.77600	4.85341	3.65096
15	-	-	-	-	-	0.50844	2.72753	4.97611	5.05539	3.14040



simulate capture histories over 15 capture occasions. The true model was therefore fully time-variant ( $\varphi, p_i$ ) and the covariate models were used to approximate this temporal structure. This procedure mimicked the analysis of real data in which the covariates are known constants and the observed capture histories are a realization of random survival and recapture processes. Given that the mean of  $\varphi$  over all occasions was approximately 0.80 in all three parameter sets, an initial release of 100 marked individuals in the first occasion was followed by the release of 20 newly marked individuals in each of the subsequent 13 occasions to maintain approximately 100 marked individuals in the population through time. The 15th occasion was a recapture event only and no newly marked individuals were released. A total of 1000 sets of capture histories were randomly generated for each of the three parameter sets.

For each set of capture histories, the  $p$ -first strategy was implemented by pairing the most general sub-model for  $\varphi$ , consisting of an intercept and all four covariates additively, with all 16 sub-models for  $p$ . From among the resulting 16 models, the model having the minimum value of  $AIC_C$  was identified, after which the corresponding sub-model for  $p$  was paired with each of the remaining 15 sub-models for  $\varphi$ . The  $\varphi$ -first strategy was conceptually identical, with the order between  $\varphi$  and  $p$  sub-model conditioning reversed. Inference under both the  $p$ - and  $\varphi$ -first strategies was therefore based upon 31 (12%) of the possible 256 sub-model combinations. Under the AC strategy, the parameters of all 256 models were estimated. The PC strategy was implemented as previously described.

The fit of estimated models was evaluated by computing the relative sum of squares (RSS):

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

where  $\theta_i$  represents an element of either  $\varphi$  or  $p$  for an individual capture occasion and  $\hat{\theta}_i$  is the corresponding estimate based on either a single model or computed by model averaging [15]. Given that there were 15 capture occasions, there were 14  $\varphi$  and 14  $p$  parameters (**Table 1**). Although all 28 parameters were theoretically estimable for models using covariates, assessing estimation performance for only 13  $\varphi$  and 13  $p$  parameters allows comparisons to be made with the fully-parameterized time-variant model used to generate capture histories, in which the last  $\varphi$  and  $p$  parameters are confounded [25,26]. The sum of RSS statistics for  $\varphi$  and  $p$  was computed to provide a measure of performance for the entire model.

For each of the four model selection strategies, the RSS for  $\varphi$ ,  $p$ , and their sum was computed using the model with the lowest  $AIC_C$ . The RSS statistics were

similarly computed for the model-averaged estimates, using the models considered under each strategy. For comparative purposes, we also report RSS statistics for the one model among all 256 having the smallest total RSS, as well as the fully time-variant model used to generate capture histories. The mean and standard deviation of RSS statistics among the 1000 capture histories with each parameter set were computed.

The importance of parameter covariates and covariate combinations was evaluated by computing cumulative  $AIC_C$  model weights [15] for each covariate and covariate combination among the models considered under each selection strategy. The mean and standard deviation of cumulative weights were computed across the 1000 replications using each parameter set. Note that the frequency with which individual covariates were included in models was equal, or balanced, under the AC strategy, while the frequency of covariate combinations was balanced under the AC and PC strategies. The simulations were conducted using version 2.14.1 of R [35] and model parameters were estimated using version 2.10 of the R package *mra* [36].

### 3. RESULTS

The AC, PC, and  $p$ - and  $\varphi$ -first evaluation strategies all yielded approximately equal means and standard deviations for the RSS statistic (**Table 2**), and the values were similar to those reported by Doherty *et al.* [34]. None of the strategies resulted in parameter estimates with a mean RSS that approached that of the single model with the smallest RSS, indicating that for each set of capture histories there was usually at least one model among all 256 models with parameter estimates closer to the true values than either the minimum  $AIC_C$  model, and closer than the model-averaged estimates. The model with the minimum RSS had an average of 1 to 1.5 more parameters than the model with minimum  $AIC_C$ . Because of the increased variance associated with estimating the large number of parameters in the fully-parameterized time-variant model ( $\varphi, p_i$ ), all strategies produced smaller mean RSS values. Model-averaging produced a modest but consistent improvement in parameter estimation.

All strategies gave the largest cumulative weights to the individual covariates most highly correlated with  $\varphi$  and  $p$  ( $x_{1\varphi}$ ,  $x_{2\varphi}$ ,  $x_{1p}$ , and  $x_{2p}$ , **Table 3**). However, covariates having low correlations with the parameters accumulated more weight than might be expected [34]. In particular, the  $\varphi$ -first strategy tended to result in relatively large cumulative weights for the covariates having small correlations with  $p$  ( $x_{3p}$  and  $x_{4p}$ ). For example, the covariate  $x_{4p}$  in Parameter set 1 had a correlation with  $p$  of only  $-0.013$ , but its cumulative weight under the  $\varphi$ -first strategy was 0.509 (**Table 3**). Similarly, the  $p$ -first strategy tended to produce relatively large cumulative weights for

**Table 2.** The mean and standard deviation (SD) of relative sum of squares (RSS) observed among 1000 simulations with each of three parameter sets for the  $\varphi$  and  $p$  parameters, the sum of the two RSS values ( $\varphi + p$ ), and the mean and standard deviation of the number of parameters in the model (No. param.). Estimation performance is summarized for the single model with the minimum RSS (Min. RSS) value, the fully-time variant model ( $\varphi p_t$ ), and each of the four model selection strategies. For each strategy, results are shown for the model with the minimum AIC<sub>C</sub> value (Min. AIC<sub>C</sub>) among the models considered, as well the model-averaged estimates (Model Ave.).

Strategy	Criterion	Mean RSS			SD of RSS			No. param.	
		$\varphi$	$p$	$\varphi + p$	$\varphi$	$p$	$\varphi + p$	Mean	SD
Parameter set 1									
	Min. RSS	0.078	0.204	0.283	0.035	0.069	0.079	8.0	1.0
	$\varphi p_t$	0.192	0.329	0.522	0.091	0.167	0.206	27.0	0.0
$p$ -first	Min. AIC <sub>C</sub>	0.113	0.255	0.369	0.065	0.097	0.117	6.5	0.9
	Model Ave.	0.105	0.252	0.357	0.056	0.092	0.109	-	-
$\varphi$ -first	Min. AIC <sub>C</sub>	0.114	0.256	0.370	0.065	0.096	0.116	6.5	1.0
	Model Ave.	0.110	0.257	0.367	0.061	0.095	0.114	-	-
PC	Min. AIC <sub>C</sub>	0.114	0.255	0.369	0.065	0.095	0.116	6.5	1.0
	Model Ave.	0.106	0.251	0.358	0.056	0.091	0.108	-	-
AC	Min. AIC <sub>C</sub>	0.114	0.255	0.369	0.065	0.095	0.116	6.5	1.0
	Model Ave.	0.106	0.252	0.358	0.056	0.091	0.107	-	-
Parameter set 2									
	Min. RSS	0.051	0.123	0.175	0.016	0.051	0.053	7.9	1.1
	$\varphi p_t$	0.100	0.197	0.298	0.052	0.105	0.121	27.0	-
$p$ -first	Min. AIC <sub>C</sub>	0.072	0.149	0.221	0.027	0.072	0.076	7.0	1.1
	Model Ave.	0.064	0.147	0.211	0.025	0.069	0.072	-	-
$\varphi$ -first	Min. AIC <sub>C</sub>	0.072	0.149	0.221	0.027	0.072	0.076	7.0	1.0
	Model Ave.	0.069	0.146	0.216	0.026	0.067	0.070	-	-
PC	Min. AIC <sub>C</sub>	0.072	0.149	0.221	0.027	0.072	0.076	7.0	1.1
	Model Ave.	0.067	0.145	0.210	0.025	0.065	0.069	-	-
AC	Min. AIC <sub>C</sub>	0.072	0.149	0.221	0.027	0.072	0.076	7.0	1.1
	Model Ave.	0.064	0.145	0.210	0.025	0.065	0.069	-	-
Parameter set 3									
	Min. RSS	0.025	0.066	0.092	0.011	0.024	0.027	7.7	1.0
	$\varphi p_t$	0.088	0.152	0.240	0.040	0.065	0.081	27.0	-
$p$ -first	Min. AIC <sub>C</sub>	0.037	0.084	0.121	0.019	0.036	0.042	6.6	1.0
	Model Ave.	0.032	0.082	0.114	0.018	0.035	0.040	-	-
$\varphi$ -first	Min. AIC <sub>C</sub>	0.037	0.084	0.121	0.018	0.036	0.041	6.6	1.0
	Model Ave.	0.035	0.082	0.116	0.017	0.034	0.039	-	-
PC	Min. AIC <sub>C</sub>	0.037	0.084	0.121	0.019	0.036	0.042	6.6	1.0
	Model Ave.	0.032	0.080	0.112	0.018	0.034	0.039	-	-
AC	Min. AIC <sub>C</sub>	0.037	0.084	0.121	0.019	0.036	0.042	6.6	1.0
	Model Ave.	0.032	0.080	0.112	0.018	0.034	0.039	-	-

**Table 3.** The mean and standard deviation (SD) of cumulative Akaike Information Criterion (AIC<sub>C</sub>) model weights for individual survival ( $x_{i\phi}$ ) and recapture ( $x_{ip}$ ) probability covariates observed among estimates from 1000 simulations with each of three parameter sets under the  $p$ -first,  $\phi$ -first, PC, and AC model-selection strategies. Bivariate correlations between the logit of the  $\phi$  and  $p$  parameters and their covariates (Corr.) are taken from [34].

Covariate	Corr.	$p$ -first		$\phi$ -first		PC		AC	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Parameter set 1									
$x_{1\phi}$	0.861	0.823	0.208	0.817	0.315	0.808	0.244	0.805	0.222
$x_{2\phi}$	0.645	0.719	0.228	0.647	0.399	0.681	0.252	0.680	0.249
$x_{3\phi}$	0.266	0.601	0.229	0.525	0.401	0.559	0.238	0.556	0.235
$x_{4\phi}$	0.052	0.490	0.168	0.312	0.341	0.432	0.169	0.432	0.167
$x_{1p}$	0.917	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
$x_{2p}$	0.660	0.989	0.076	0.975	0.076	0.969	0.092	0.966	0.091
$x_{3p}$	0.325	0.195	0.310	0.508	0.146	0.367	0.136	0.367	0.136
$x_{4p}$	-0.013	0.199	0.314	0.509	0.147	0.370	0.139	0.370	0.138
Parameter set 2									
$x_{1\phi}$	0.864	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
$x_{2\phi}$	0.484	0.753	0.217	0.737	0.362	0.708	0.239	0.706	0.236
$x_{3\phi}$	0.262	0.694	0.228	0.641	0.400	0.641	0.248	0.639	0.246
$x_{4\phi}$	0.058	0.531	0.170	0.293	0.352	0.420	0.168	0.420	0.167
$x_{1p}$	0.900	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
$x_{2p}$	0.491	0.914	0.195	0.900	0.149	0.891	0.159	0.899	0.158
$x_{3p}$	0.228	0.248	0.304	0.484	0.159	0.380	0.153	0.379	0.152
$x_{4p}$	0.052	0.282	0.327	0.499	0.168	0.396	0.161	0.396	0.161
Parameter set 3									
$x_{1\phi}$	0.865	0.986	0.045	0.993	0.044	0.986	0.053	0.984	0.053
$x_{2\phi}$	0.706	0.722	0.229	0.704	0.362	0.693	0.243	0.692	0.241
$x_{3\phi}$	-0.005	0.452	0.155	0.236	0.289	0.376	0.146	0.376	0.145
$x_{4\phi}$	0.063	0.521	0.205	0.381	0.376	0.463	0.211	0.462	0.210
$x_{1p}$	0.817	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.001
$x_{2p}$	0.470	0.972	0.130	0.962	0.093	0.953	0.116	0.950	0.115
$x_{3p}$	0.225	0.257	0.353	0.535	0.175	0.408	0.172	0.408	0.171
$x_{4p}$	0.076	0.260	0.355	0.532	0.174	0.406	0.170	0.405	0.170

the covariates having small correlations with  $\phi$  ( $x_{3\phi}$  and  $x_{4\phi}$ ). The AC and PC strategies produced cumulative weights for these covariates that were intermediate between those of the  $p$ - and  $\phi$ -first strategies. However, note that the use of individual covariates was not balanced under the  $p$ - and  $\phi$ -first strategies (*see DISCUSSION*).

Substantial differences in the standard deviations of the cumulative weights for individual covariates were also observed among the strategies. For example, consider the  $p$ -first results for Parameter set 1 (**Table 3**). The standard deviations of the cumulative weights for  $x_{3p}$  and

$x_{4p}$ , which both have low correlations with  $p$ , were more than double the values observed under the other strategies. Similar results were obtained for  $x_{3\phi}$  and  $x_{4\phi}$  under the  $\phi$ -first strategy, and the results were consistent across the three parameter sets. Covariates that might be considered unimportant on the basis of pairwise correlations accumulated considerable AIC<sub>C</sub> weight if they were included in the first sub-model selected and so were also included in the subsequent 15 models, but accumulated reduced weight otherwise, resulting in elevated variability.

The difference among the strategies was perhaps most

clearly revealed by the cumulative model weights for combinations of covariates, which represent model structure more effectively than weights associated with individual covariates. The results were similar for all three parameter sets, so only the results for Parameter set 1 are

shown (**Table 4**). Under the  $p$ -first strategy, the most complex sub-model for  $\varphi$  had a mean cumulative weight of 0.204, approximately twice its mean weight of 0.106 observed under the AC strategy.

The standard deviation of this  $\varphi$  sub-model was also

**Table 4.** The mean and standard deviation (SD) of cumulative Akaike Information Criterion (AIC<sub>C</sub>) model weights for combinations of survival ( $\varphi$ ) and recapture ( $p$ ) probability covariates observed among estimates from 1,000 simulations with parameter set 1 under the  $p$ -first,  $\varphi$ -first, PC, and AC model-selection strategies.

Covariates	$p$ -first		$\varphi$ -first		PC		AC	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i><math>\varphi</math> covariates</i>								
-	0.001	0.006	0.001	0.004	0.001	0.006	0.001	0.006
1	0.062	0.084	0.094	0.232	0.068	0.090	0.069	0.089
2	0.056	0.083	0.087	0.222	0.061	0.089	0.061	0.088
3	0.001	0.004	0.000	0.003	0.001	0.005	0.001	0.005
4	0.001	0.004	0.000	0.002	0.001	0.004	0.001	0.004
1+2	0.110	0.094	0.152	0.275	0.122	0.101	0.121	0.100
1+3	0.096	0.096	0.120	0.255	0.102	0.105	0.102	0.104
1+4	0.038	0.053	0.026	0.101	0.042	0.060	0.043	0.058
2+3	0.036	0.054	0.026	0.100	0.039	0.058	0.039	0.057
2+4	0.053	0.072	0.056	0.168	0.058	0.080	0.059	0.078
3+4	0.001	0.003	0.000	0.001	0.000	0.003	0.001	0.003
1+2+3	0.153	0.100	0.209	0.318	0.175	0.125	0.174	0.122
1+2+4	0.078	0.061	0.061	0.158	0.089	0.070	0.088	0.070
1+3+4	0.087	0.108	0.112	0.257	0.104	0.135	0.103	0.134
2+3+4	0.029	0.040	0.013	0.048	0.032	0.045	0.032	0.044
1+2+3+4	0.204	0.130	0.045	0.108	0.107	0.081	0.106	0.080
<i><math>p</math> covariates</i>								
-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.005	0.054	0.010	0.032	0.012	0.038	0.013	0.038
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1+2	0.655	0.386	0.304	0.127	0.389	0.133	0.388	0.132
1+3	0.001	0.005	0.006	0.018	0.007	0.022	0.008	0.021
1+4	0.003	0.036	0.006	0.021	0.007	0.025	0.008	0.025
2+3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2+4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3+4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1+2+3	0.139	0.272	0.172	0.075	0.222	0.096	0.221	0.096
1+2+4	0.141	0.274	0.172	0.074	0.225	0.098	0.224	0.097
1+3+4	0.002	0.019	0.004	0.014	0.005	0.018	0.005	0.018
2+3+4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1+2+3+4	0.053	0.154	0.327	0.144	0.133	0.089	0.133	0.089

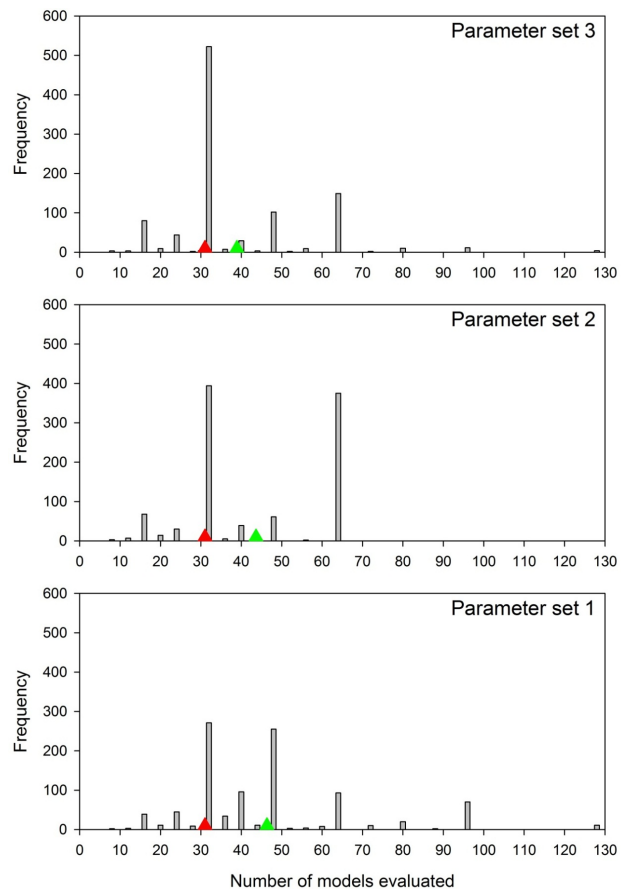


elevated under the  $p$ -first strategy; 0.130 versus 0.080. Sub-models for  $\varphi$  having fewer covariates therefore tended to have smaller mean weights and slightly reduced standard deviations under the  $p$ -first strategy. Conversely, under the  $p$ -first strategy, the more complex sub-model for  $p$  tended to accumulate less weight, sub-models with fewer covariates (especially  $x_{1p} + x_{2p}$ ) tended to accumulate more weight, and the cumulative weights of all sub-models were more variable than under the AC strategy. The results were symmetric for the  $\varphi$ -first strategy (Table 4). The results for the PC and AC strategies were nearly identical in all cases (Table 4).

The numbers of models evaluated under the  $\varphi$ -first,  $p$ -first, and AC strategies are known in advance, while the number of models evaluated under the PC strategy cannot be predetermined. The frequency distribution of the number of models evaluated by the PC strategy for each of the three parameter sets is presented in Figure 1. The mean number of models evaluated under the PC strategy was 150%, 141%, and 126% of the number of models evaluated under the  $\varphi$ -first and  $p$ -first strategies for parameter sets 1, 2, and 3, respectively. However, in 11%, 12%, and 15% of the simulations, the number of models evaluated under the PC strategy was less than the number evaluated under the  $\varphi$ -first and  $p$ -first strategies. The maximum number of models evaluated under the PC strategy was 128, 64, and 128, which were either 50% or 25% of the number of models evaluated by the AC strategy (256). Note that these statistics do not include the initial models whose results were used to identify plausible sub-models.

#### 4. DISCUSSION

The PC strategy has several attractive characteristics. Under the  $p$ - and  $\varphi$ -first strategies, one initially conditions upon a complex sub-model for either  $p$  or  $\varphi$ , while evaluating candidate sub-models for the other parameter class. The estimation results from these models are carried forward and constitute over half of the models that are evaluated (16 of 31 models in this investigation). The cumulative weights in Tables 2 and 3 suggest that these stepwise procedures tend to inflate the apparent importance of complex covariate models for one parameter class and, consequently, overly simplistic covariate models for the other parameter class. This phenomenon constitutes an information “leak” between sub-models [32]. Neither strategy produces an accurate assessment of covariate importance for either  $\varphi$  or  $p$ . While flexible sub-models for either  $p$  or  $\varphi$  are also conditioned upon in the first step of the PC strategy, the only results from the initial modeling step that are carried forward are lists of sub-models for subsequent evaluation, *i.e.*, the initial parameter estimates are discarded. Consequently, results of models evaluated during the search through the model



**Figure 1.** The frequency distribution of the number of models on which inference was based under the PC strategy using each of the three parameters sets, with the mean number of models indicated by a green triangle. The number of models under the  $\varphi$ -first and  $p$ -first strategies (31) is indicated with a red triangle, while the number of models under the AC strategy (256) is not shown.

space are disassociated from model results on which inference is based, and the selection of a flexible sub-model in the initial modeling step has little influence on the ultimate parameter estimates and covariate weights obtained under the PC strategy. Unlike the  $p$ - and  $\varphi$ -first strategies, the PC strategy is symmetric with respect to  $p$  and  $\varphi$  and the order in which they are considered is immaterial. Finally, the PC strategy explores the potential model space more effectively than the  $p$ - and  $\varphi$ -first strategies by considering all models that provide a comparatively good approximation to the data, conditioned on the candidate model set, leading to an improved assessment of the importance of covariates and covariate combinations, and its numerical performance was indistinguishable from that of the AC strategy.

One potential disadvantage of the PC strategy is that the number of models to be evaluated cannot be predetermined, and will vary as a function of the observed capture histories and the candidate sub-models. Logically,

one can expect the number of models evaluated to increase with the number of competitive sub-models, especially if there are numerous competitive sub-models for both  $p$  and  $\varphi$ . However, note that the PC strategy can be modified in straightforward fashion to influence the number of models evaluated. For example, the threshold  $AIC_C$  model weight of 0.01 used in this investigation could be increased, which would reduce the number of models evaluated. In addition, a threshold based on  $\Delta AIC_C$  could be implemented. In any case, achieving performance equivalent to that of the AC strategy while estimating the parameters of far fewer models should outweigh concern about how many models need to be evaluated in many investigations, especially if parameter estimation can be performed efficiently, as is possible in the R programming environment [35,36].

The equivalent performance of all strategies with respect to parameter estimation, as measured by RSS, is intriguing. We suspect that this is a consequence of working with several additive covariate combinations that provide approximately equal explanatory power. However, based on our experience modeling real data, it is not uncommon for models with different covariate structures to produce parameter estimates with non-trivial differences but have similar model weights. The information “leak” between sub-models [32] that is apparent in the results of the  $p$ - and  $\varphi$ -first strategies is indicative that this may have occurred to some degree in the simulation. Under such conditions, differences among the strategies with respect to parameter estimation are more likely to exist and model-averaging is likely to provide greater benefits than our results suggest. In any case, we caution against the potentially naïve interpretation of our results that use of either the  $p$ - or  $\varphi$ -first strategies is adequate if one is primarily interested in the estimation of natural parameters, rather than the importance of covariates or covariate structure.

The issue of covariate balance in a model set is critical when using covariate weights to assess the explanatory utility of covariates. Covariates that are included in a large number of models can accumulate substantial weight even if they are not truly informative. Conversely, a covariate with meaningful explanatory value could conceivably accumulate lower weight than it merits if included in few models, although those models should tend to have relatively high weights. Of the model evaluation strategies considered in this investigation, only the AC strategy was balanced for individual covariates. Various approaches to adjust for unequal covariate balance within a model set have been implemented [34, 37], although no such adjustments were made in this investigation (Table 3) because differences among the strategies were being evaluated. Even if such methods correctly adjust for the number of times a covariate is

used, our view is that the use of model weights to assess the importance of individual covariates is overemphasized. Covariate combinations define model structure and covariates are rarely uncorrelated, hence model structure is more effectively assessed by examining cumulative weights associated with unique covariate combinations [38]. Note that the subset of models evaluated under the PC strategy is naturally balanced for covariate combinations.

We would like to reemphasize the importance of deliberately formulating a candidate model set with careful consideration of prior information, expert opinion, ecological relevance, and research objectives or hypotheses [13,15,16,18]. Including every conceivable model and using an automated or semi-automated strategy to search the model space is data or model dredging [15], and can lead to spurious results even when using model averaging. If such an approach is implemented, the utility of the resulting model or parameter estimates should be assessed using cross validation or the collection of independent data. However, even when model sets are constructed with great care, the number of models included can still be excessively large. In cases with an unfeasibly large number of models to evaluate, an objective algorithm for searching the model space, such as the PC strategy, is likely to prove useful.

Model selection and the evaluation of model structure will continue to be fundamental aspects of mark-recapture modeling. The PC strategy offers substantial advantages compared to other strategies in common use, producing parameter estimates and cumulative covariate weights that are nearly indistinguishable from those produced by the AC strategy, without substantially increasing the average number of models evaluated in comparison to the  $p$ - and  $\varphi$ -first strategies. Although the performance of the PC strategy was investigated using CJS mark-recapture models, the conceptual approach is easily extensible to other classes of models in which the global model can be viewed as a combination of component sub-models, such as multistate CJS models. An investigation of the performance of the PC strategy with more complex models would be informative.

## 5. ACKNOWLEDGEMENTS

We would like to offer our sincere gratitude to Paul Doherty and Gary White for providing the parameter and covariate values on which the simulations were based, which strengthened the comparative basis of this investigation. We also thank Mark Udevitz and Karen Oakley, US Geological Survey, Alaska Science Center, and anonymous reviewers for constructive comments on drafts of the manuscript. Any mention of trade names is for descriptive purposes only and does not imply endorsement by the US Government, Western Ecosystems Technology, Inc., or Polar Bears International.

## REFERENCES

- [1] Kuparinen, A., Alho, J.S., Olin, M. and Lehtonen, H. (2012) Estimation of northern pike population sizes via mark-recapture monitoring. *Fisheries Management and Ecology*, **19**, 323-332. [doi:10.1111/j.1365-2400.2012.00846.x](https://doi.org/10.1111/j.1365-2400.2012.00846.x)
- [2] Harris, R.B., Winnie, J., Amish, S.J., Beja-Pereira, A., Godinho, R., Costa, V. and Luikart, G. (2010) Argali abundance in the Afghan Pamir using capture-recapture modeling from fecal DNA. *Journal of Wildlife Management*, **74**, 668-677. [doi:10.2193/2009-292](https://doi.org/10.2193/2009-292)
- [3] Sandercock, B.K. and Jaramillo, A. (2002) Annual survival rates of wintering sparrows: Assessing demographic consequences of migration. *The Auk*, **119**, 149-165.
- [4] Duriez, O., Sæther, S.A., Ens, B.J., Choquet, R., Pradel, R., Lambeck, R.H.D. and Klaassen, M. (2009) Estimating survival and movements using both live and dead recoveries: A case study of oystercatchers confronted with habitat change. *Journal of Applied Ecology*, **46**, 144-153. [doi:10.1111/j.1365-2664.2008.01592.x](https://doi.org/10.1111/j.1365-2664.2008.01592.x)
- [5] Doherty, P.F., Nichols, J.D., Tautin, J., Voelzer, J.F., Smith, G.W., Benning, D.S., Bentley, V.R., Bidwell, J.K., Bollinger, K.S., Brazda, A.R., Buelna, E.K., Goldsberry, J.R., King, R.J., Roetker, F.H., Solberg, J.W., Thorpe, P.P. and Wortham, J.S. (2002) Sources of variation in breeding-ground fidelity of mallards (*Anas platyrhynchos*). *Behavioral Ecology*, **13**, 543-550. [doi:10.1093/beheco/13.4.543](https://doi.org/10.1093/beheco/13.4.543)
- [6] Punt, A.E., Buckworth, R.C., Dichmont, C.M. and Ye, Y. (2009) Performance of methods for estimating size-transition matrices using tag-recapture data. *Marine and Freshwater Research*, **60**, 168-182. [doi:10.1071/MF08217](https://doi.org/10.1071/MF08217)
- [7] Conn, P.B. and Cooch, E.G. (2009) Multistate capture-recapture analysis under imperfect state observation: An application to disease models. *Journal of Applied Ecology*, **46**, 486-492. [doi:10.1111/j.1365-2664.2008.01597.x](https://doi.org/10.1111/j.1365-2664.2008.01597.x)
- [8] Schwarz, C.J., Bailey, R.E., Irvine, J.R. and Dalziel, F.C. (1993) Estimating salmon spawning escapement using capture-recapture methods. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 1181-1197. [doi:10.1139/f93-135](https://doi.org/10.1139/f93-135)
- [9] O'Hara, R.B., Lampila, S. and Orell, M. (2009) Estimation of rates of births, deaths, and immigration from mark-recapture data. *Biometrics*, **65**, 275-281. [doi:10.1111/j.1541-0420.2008.01048.x](https://doi.org/10.1111/j.1541-0420.2008.01048.x)
- [10] Conn, P.B., Gorgone, A.M., Jugovich, A.R., Byrd, B.L. and Hansen, L.J. (2011) Accounting for transients when estimating abundance of bottlenose dolphins in Choctawhatchee Bay, Florida. *The Journal of Wildlife Management*, **75**, 569-579. [doi:10.1002/jwmg.94](https://doi.org/10.1002/jwmg.94)
- [11] Fearnbach, H., Durban, J., Parsons, K. and Claridge, D. (2012) Photographic mark-recapture analysis of local dynamics within an open population of dolphins. *Ecological Applications*, **22**, 1689-1700.
- [12] White, G.C. and Burnham, K.P. (1999) Program MARK: Survival estimation from populations of marked animals. *Bird Study*, **46**, S120-S139. [doi:10.1080/00063659909477239](https://doi.org/10.1080/00063659909477239)
- [13] Steidl, R.J. (2007) Limits of data analysis in scientific inference: Reply to Sleep et al. *Journal of Wildlife Management*, **71**, 2122-2124. [doi:10.2193/2007-187](https://doi.org/10.2193/2007-187)
- [14] Shaffer, T.L. and Johnson, D.H. (2008) Ways of learning: Observational studies versus experiments. *Journal of Wildlife Management*, **72**, 4-13. [doi:10.2193/2007-293](https://doi.org/10.2193/2007-293)
- [15] Burnham, K.P. and Anderson, D.R. (2001) Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, **28**, 111-119. [doi:10.1071/WR99107](https://doi.org/10.1071/WR99107)
- [16] Lukacs, P.M., Burnham, K.P. and Anderson, D. (2010) Model selection bias and Freedman's paradox. *Annals of the Institute of Statistical Mathematics*, **62**, 117-125. [doi:10.1007/s10463-009-0234-4](https://doi.org/10.1007/s10463-009-0234-4)
- [17] Sleep, D.J.H., Drever, M.C. and Nudds, T.D. (2007) Statistical versus biological hypothesis testing: Response to Steidl. *Journal of Wildlife Management*, **71**, 2120-2121. [doi:10.2193/2007-140](https://doi.org/10.2193/2007-140)
- [18] Sisson, S.A. and Fan, Y. (2009) Towards automating model selection for a mark-recapture-recovery analysis. *Journal of the Royal Statistical Society: Series C*, **58**, 247-266. [doi:10.1111/j.1467-9876.2008.00656.x](https://doi.org/10.1111/j.1467-9876.2008.00656.x)
- [19] Stephens, P.A., Buskirk, S.W., Hayward, G.D. and Martinez Del Rio, C. (2005) Information theory and hypothesis testing: A call for pluralism. *Journal of Applied Ecology*, **42**, 4-12. [doi:10.1111/j.1365-2664.2005.01002.x](https://doi.org/10.1111/j.1365-2664.2005.01002.x)
- [20] Jakeman, A.J., Letcher, R.A. and Norton, J.P. (2006) Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, **21**, 602-614. [doi:10.1016/j.envsoft.2006.01.004](https://doi.org/10.1016/j.envsoft.2006.01.004)
- [21] Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: The distinction between—and reconciliation of—"predictive" and "explanatory" models. *Biodiversity and Conservation*, **9**, 655-671. [doi:10.1023/A:1008985925162](https://doi.org/10.1023/A:1008985925162)
- [22] Hunter, C.M., Caswell, H., Runge, M.C., Regehr, E.V., Amstrup, S.C. and Stirling, I. (2010) Climate change threatens polar bear populations: A stochastic demographic analysis. *Ecology*, **91**, 2883-2897. [doi:10.1890/09-1641.1](https://doi.org/10.1890/09-1641.1)
- [23] McCrea, R.S. and Morgan, B.J.T. (2011) Multistate mark-recapture model selection using score tests. *Biometrics*, **67**, 234-241. [doi:10.1111/j.1541-0420.2010.01421.x](https://doi.org/10.1111/j.1541-0420.2010.01421.x)
- [24] King, R. and Brooks, S.P. (2008) On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, **64**, 816-824. [doi:10.1111/j.1541-0420.2007.00938.x](https://doi.org/10.1111/j.1541-0420.2007.00938.x)
- [25] Lebreton, J.D., Burnham, K.P., Clobert, J. and Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals—A unified approach with case-studies. *Ecological Monographs*, **62**, 67-118. [doi:10.2307/2937171](https://doi.org/10.2307/2937171)
- [26] Amstrup, S.C., McDonald, T.L. and Manly, B.F.J. (2005). Handbook of capture-recapture analysis. Princeton University Press, Princeton.
- [27] Regehr, E.V., Hunter, C.M., Caswell, H., Amstrup, S.C.

- and Stirling, I. (2010) Survival and breeding of polar bears in the southern Beaufort Sea in relation to sea ice. *Journal of Animal Ecology*, **79**, 117-127. [doi:10.1111/j.1365-2656.2009.01603.x](https://doi.org/10.1111/j.1365-2656.2009.01603.x)
- [28] Stirling, I., McDonald, T.L., Richardson, E.S., Regehr, E.V. and Amstrup, S.C. (2011) Polar bear population status in the northern Beaufort Sea, Canada, 1971-2006. *Ecological Applications*, **21**, 859-876. [doi:10.1890/10-0849.1](https://doi.org/10.1890/10-0849.1)
- [29] Chaloupka, M., and Limpus, C. (2005) Estimates of sex- and age-class-specific survival probabilities for a southern Great Barrier Reef green sea turtle population. *Marine Biology*, **146**, 1251-1261. [doi:10.1007/s00227-004-1512-6](https://doi.org/10.1007/s00227-004-1512-6)
- [30] Vögeli, M., Laiolo, P., Serrano, D. and Tella, J.L. (2008) Who are we sampling? Apparent survival differs between methods in a secretive species. *Oikos*, **117**, 1816-1823. [doi:10.1111/j.1600-0706.2008.17225.x](https://doi.org/10.1111/j.1600-0706.2008.17225.x)
- [31] Marucco, F., Pletscher, D.H., Boitani, L., Schwartz, M.K., Pilgrim, K.L. and Lebreton, J. (2009) Wolf survival and population trend using non-invasive capture-recapture techniques in the western Alps. *Journal of Applied Ecology*, **46**, 1003-1010. [doi:10.1111/j.1365-2664.2009.01696.x](https://doi.org/10.1111/j.1365-2664.2009.01696.x)
- [32] Catchpole, E.A., Fan, Y., Morgan, B.J.T., Clutton-Brock, T. and Coulson, T. (2004) Sexual dimorphism, survival and dispersal in red deer. *Journal of Agricultural, Biological, and Environmental Statistics*, **9**, 1-26. [doi:10.1198/1085711043172](https://doi.org/10.1198/1085711043172)
- [33] Bromaghin, J.F., Gates, K.S. and Palmer, D.E. (2010) A likelihood framework for joint estimation of salmon abundance and migratory timing using telemetric mark-recapture. *North American Journal of Fisheries Management*, **30**, 1385-1394. [doi:10.1577/M10-065.1](https://doi.org/10.1577/M10-065.1)
- [34] Doherty, P., White, G. and Burnham, K. (2012) Comparison of model building and selection strategies. *Journal of Ornithology*, **152**, 317-323. [doi:10.1007/s10336-010-0598-5](https://doi.org/10.1007/s10336-010-0598-5)
- [35] RDCT (R Development Core Team) (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://cran.r-project.org/>
- [36] McDonald, T. (2012) mra: Analysis of mark-recapture data. R package version 2.10. <http://cran.r-project.org/web/packages/mra/index.html>
- [37] Thinh, V.T., Doherty, P.F., Bui, T.H. and Huyvaert, K.P. (2012) Road crossing by birds in a tropical forest in northern Vietnam. *The Condor*, **114**, 639-644. [doi:10.1525/cond.2012.100199](https://doi.org/10.1525/cond.2012.100199)
- [38] Pyne, M.I., Byrne, K.M., Holfelder, K.A., McManus, L., Buhnerkempe, M., Burch, N., Childers, E., Hamilton, S., Schroeder, G. and Doherty Jr., P.F. (2010) Survival and breeding transitions for a reintroduced bison population: A multistate approach. *Journal of Wildlife Management*, **74**, 1463-1471.