

ORIGINAL RESEARCH

Bayesian model selection for spatial capture–recapture models

Soumen Dey  | Mohan Delampady | Arjun M. Gopaldaswamy

Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore, India

CorrespondenceArjun M. Gopaldaswamy, Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore 560059, India.
Email: arjungswamy@gmail.com**Funding information**

Indian Statistical Institute; Wildlife Conservation Society, New York

Abstract

A vast amount of ecological knowledge generated over the past two decades has hinged upon the ability of model selection methods to discriminate among various ecological hypotheses. The last decade has seen the rise of Bayesian hierarchical models in ecology. Consequently, commonly used tools, such as the AIC, become largely inapplicable and there appears to be no consensus about a particular model selection tool that can be universally applied. We focus on a specific class of competing Bayesian spatial capture–recapture (SCR) models and apply and evaluate some of the recommended Bayesian model selection tools: (1) Bayes Factor—using (a) Gelfand–Dey and (b) harmonic mean methods, (2) Deviance Information Criterion (DIC), (3) Watanabe–Akaike's Information Criterion (WAIC) and (4) posterior predictive loss criterion. In all, we evaluate 25 variants of model selection tools in our study. We evaluate these model selection tools from the standpoint of selecting the “true” model and parameter estimation. In all, we generate 120 simulated data sets using the true model and assess the frequency with which the true model is selected and how well the tool estimates N (population size), a parameter of much importance to ecologists. We find that when information content is low in the data, no particular model selection tool can be recommended to help realize, simultaneously, both the goals of model selection and parameter estimation. But, in general (when we consider both the objectives together), we recommend the use of our application of the Bayes Factor (Gelfand–Dey with MAP approximation) for Bayesian SCR models. Our study highlights the point that although new model selection tools are emerging (e.g., WAIC) in the applied statistics literature, those tools based on sound theory even under approximation may still perform much better.

KEYWORDS

Bayes factors, Bayesian inference, DIC, hierarchical models, posterior predictive loss, WAIC

1 | INTRODUCTION

Following the highly influential text, Burnham and Anderson (2002), on model selection, ecologists and conservation biologists have drastically shifted their inferential practice from the “hypothesis testing” approach to the more appropriate “hypothesis discrimination”

approach (Johnson & Omland, 2004). To date, Burnham and Anderson (2002) has been cited 44,168 times (as on January 3, 2019—Google Scholar), demonstrating the impact of this text. One may argue that this contribution has helped increase the pace of growth in ecological knowledge because it has paved the way for researchers to draw inferences more robustly because of the ability

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

to now assess the influence of various competing a priori hypotheses (models) without altering the study question to suit the restrictive hypothesis testing paradigm (Bolker, 2008).

Since a large amount of ecological data are based on field observations, it called for ecologists to take the approach of “detectives” rather than “hypothetico-deductive” scientists by formulating models using likelihood functions to confront various a priori hypotheses using observational data (Hilborn & Mangel, 1997). And by maximizing the likelihood and using a model selection tool, such as the Akaike’s information criterion (Burnham & Anderson, 2002), researchers found a way to place increased faith on models favored by such criteria. Thus, a vast amount of ecological knowledge generated has relied on the robustness of such model selection tools in accurately discriminating hypotheses.

Recently, there has been an increased use of hierarchical models in ecology since they appear to address two important issues: (a) ecological scales are naturally hierarchical in structure and (b) hierarchical models form a natural way of incorporating the observation process (Royle & Dorazio, 2008). With powerful tools such as the MCMC, it is now possible to confront complex ecological models with data in a Bayesian inferential framework (Bolker, 2008). However, it remains unclear as to how to discriminate among competing hypotheses (models) because popular model selection criteria (such as AIC, BIC, or DIC; Burnham & Anderson, 2002) are not easy to apply or work poorly for complex hierarchical models (Millar, 2009).

It is well known that, asymptotically, the Bayes factor is the preferred model selection tool due to its consistency property, that is, to identify the true data-generating model if (and only if) the true data-generating model is included in the model set and if the data tend to the limit of infinite informativeness (Ghosh, Delampady, & Samanta, 2006; Robert, 2007). However, this property holds only under certain regularity conditions that are often difficult to verify for complex models (Berger, Ghosh, & Mukhopadhyay, 2003; Dass & Lee, 2004; Ghosh & Samanta, 2001). More prominently, there is vast literature expressing the difficulties in computing the marginal likelihood in applied statistical problems (Chan & Eisenstat, 2015; Wang, 2018). Hence, it becomes necessary to also consider alternatives to the Bayes factor or to find novel ways of applying them in practice.

Recently, Hooten and Hobbs (2015) summarized a wide array of Bayesian model selection methods that are available to ecologists. However, the generality of the recommendations provided by them remains unknown. Given such innate difficulties involved in discovering the “ideal” model selection tool both from the standpoint of theory and its application to a broad class of models, it appears to be prudent to explore the model selection issue by conditioning, at least, on a particular class of models.

Here, we evaluate various Bayesian model selection tools on a class of Bayesian spatial capture–recapture (SCR) models that are now used frequently for animal density estimation (Royle, Chandler, Sollmann, & Gardner, 2013). Although, previously, Goldberg et al. (2015) has attempted to apply the Bayes Factor (Gelfand–Dey estimator) in an abundance estimation problem for leopards (*Panthera*

pardus), their approach of computing the ratio term in the estimator seems inaccurate in the context of how the denominator has to be computed according to Gelfand and Dey (1994). Thus, we evaluate various Bayesian model selection tools by: (a) defining a class of competing models (in our case, these include the model developed in Dey, Delampady, Karanth, and Gopalaswamy (2019) along with simplified alternatives) that vary both in terms of structural and model complexity (b) simulating data sets from a “true” model (c) practically implementing a variety of Bayesian model selection tools, and in specific cases, also proposing alternatives previously not defined (d) assessing the efficacy of these implementations from the standpoint of model selection and parameter estimation and (e) providing recommendations to practitioners based on our results.

2 | METHODS

We describe here the sampling design and development of the competing models in the candidate model set.

2.1 | The candidate model set

Capture–recapture surveys are conducted by placing an array of detectors (e.g., camera traps, hair snares) to sample the species of interest within a bounded region over a fixed period of time. As an extension, spatial capture–recapture (SCR) models draw inference on the spatial distribution of animals using their spatial locations from the recorded capture–recapture samples (Royle et al., 2013).

In photographic capture–recapture surveys, an array of camera traps are placed over the study area. Each camera trap station consists of two cameras placed opposite to one another to photograph the flanks of animals passing by. Naturally marked species can be identified by the unique patterns on their flanks. However, an animal passing through a camera trap station does not necessarily result in identifiable flank photographs from both the sides. This is because there are many known and unknown factors that can influence camera firing rates. Such a detection process results in uncertain or partially identified individuals in the spatial capture–recapture sample and thus provided the motivation for the development of Dey et al. (2019). We summarize the description of the model below.

2.1.1 | Sampling situation

The notation used in this article is described in Tables 1 and 2. However, we describe a few variables and parameters for ease in the model description below. Consider a capture–recapture survey of a species with naturally marked individuals in which two detectors are collocated at J trap stations (within a bounded geographic region $\mathcal{V} \subset \mathbb{R}^2$) and kept active for K sampling occasions. An individual can be completely identified if both the detectors record the individual simultaneously at least once during the course of study (Royle, 2015). We assume that each detector captures some mutually exclusive attributes of an individual. These capture outcomes

TABLE 1 Notations of variables and parameters used in this article

Variables and parameters	Definition
\mathcal{V}	A bounded geographic region of scientific or operational relevance where a population of individuals of certain species reside
$N \sim \text{Binomial}(M, \psi)$	Population size of the superpopulation, that is, the number of individuals within \mathcal{V}
M	Maximum number of individuals within the state space \mathcal{V} This is a fixed quantity defined by the investigator
ψ	Proportion of individuals that are real and present within \mathcal{V}
θ	Probability that an individual is male
J	Number of trap stations in \mathcal{V}
K	Number of sampling occasions
R	Maximum permissible value of movement range for each individual during the survey
ω_0	Baseline trap entry probability in the models M_1 and M_2 , that is, probability that an individual passes through a trap station assuming its center of activity is also located at that trap station
p_0	Baseline detection probability in the models M_2 and M_4 , that is, probability that an individual is detected by a detector assuming its center of activity is also located at that trap station
σ	σ measures the spatial extent of movement around individual activity center. $\sigma = \sigma_m$ for male individuals, $\sigma = \sigma_f$ for female individuals
$d_{ij} = d(\mathbf{s}_i, \mathbf{x}_j) = \ \mathbf{s}_i - \mathbf{x}_j\ $	Euclidean distance between points \mathbf{s}_i and \mathbf{x}_j
$\eta_j(\mathbf{s}_i, u_i) = \omega_0 \exp\left(-\frac{d(\mathbf{s}_i, \mathbf{x}_j)^2}{2\sigma(u_i)^2}\right)$	Probability that an individual i passes through a trap station \mathbf{x}_j on some occasion k and σ is modeled as a function of individual covariate on sex category u_i
$\eta_j(\mathbf{s}_i) = \omega_0 \exp\left(-\frac{d(\mathbf{s}_i, \mathbf{x}_j)^2}{2\sigma^2}\right)$	Probability that an individual i passes through a trap station \mathbf{x}_j on some occasion k
ϕ	Probability that an individual i is detected by a detector on some occasion k given that it is present at that trap
Notations pertaining to model selection tools	
$\hat{m}_{\text{GD}}(\mathbf{Y})$	Gelfand-Dey estimator of the marginal likelihood of data $m(\mathbf{Y})$
$\hat{m}_{\text{HM}}(\mathbf{Y})$	Harmonic mean estimator of the marginal likelihood of data $m(\mathbf{Y})$
p_{DIC}	Correction term for bias due to overfitting in DIC criterion
p_{WAIC}	Correction term for bias due to overfitting in WAIC criterion
D_∞	Posterior predictive loss criterion

Note: Bold symbols represent collections (vectors).

are recorded as binary observations $y_{ijk}^{(1)}$ and $y_{ijk}^{(2)}$ for an individual i at trap station \mathbf{x}_j on sampling occasion k corresponding to detectors 1 and 2, respectively. The paired Bernoulli outcomes $y_{ijk} = (y_{ijk}^{(1)}, y_{ijk}^{(2)})$ give rise to bilateral spatial capture–recapture data for each individual i at location \mathbf{x}_j on occasion k . The array of a bilateral capture history for an individual i is denoted by $\mathbf{Y}_{i,\text{obs}} = (\mathbf{Y}_{i,\text{obs}}^{(1)}, \mathbf{Y}_{i,\text{obs}}^{(2)}) = ((y_{ijk}^{(1)}, y_{ijk}^{(2)}))_{j,k}$, which is of dimension $2 \times J \times K$. In Example 2.1, we provide an example of a sample data set coming out of a spatial capture–recapture survey with two detectors deployed at each station.

Example 2.1 Suppose a capture–recapture survey is conducted where a pair of detectors (1 and 2) are deployed at each of the 3 ($= J$) trap stations and kept active for 4 ($= K$) sampling occasions. Two individuals get fully identified based on their obtained capture histories (captured in both cameras at least once during the survey). The capture history for each of these two fully identified individuals is of dimension $2 \times 3 \times 4$. The detection histories are tabulated in Table 3. Here, individual 1 is fully identified owing to the capture event at trap 2 on occasion 4. Individual 1 is also fully identified as it is captured

at trap 2 on occasion 4. We note that in practical applications, full identity of individuals is obtained only when there are metadata, such as time of capture, to ascertain simultaneous captures. Due to the absence of simultaneous capture events in the detection histories of the partially identified individuals, we are uncertain about whether these histories correspond to two different individuals or to the same individual.

2.1.2 | Model likelihoods

We have considered four models, denoted by M_1 , M_2 , M_3 , and M_4 , and the model likelihoods of the corresponding models follow in Equations (1–4). Each of the four models M_1 – M_4 are SECR models for partially identified capture–recapture history. The basic difference between M_3 and M_4 is that M_4 does not explain the latent hierarchy of the two events: trap entry of an animal and subsequent detection conditional on trap entry, whereas the former model M_3 does. M_1 uses the covariate information on sex category and allows the movement scale parameter (σ in our case) of M_3 to be gender specific. The

TABLE 2 Notations of latent variables and data used in this article

Latent variables	Definition
S	Locations of the activity centers of N animals within \mathcal{V}
$s_i = (s_{i1}, s_{i2})'$	Location of individual i 's activity center.
$Z = (z_1, z_2, \dots, z_M)'$	A vector of Bernoulli variables, $z_i = 1$ if individual i is present
$u = (u_1, \dots, u_M)'$	A vector of Bernoulli variables, $u_i = 1$ if individual i is male in the population and $u_i = 0$ if it is a female
$u_0(C, U)$	Vector of "missing" binary observations on sexes of the list of M individuals
$L = (L_1, L_2, \dots, L_M)'$	Each L_i takes value in $\{1, 2, \dots, M\}$ and denotes the true index of i th detector 2 individual
Data	Definition
$x_j = (x_{j1}, x_{j2})'$	Location of j th trap station for detectors
$u_{\text{obs}}(C, U)$	Vector of "recorded" binary observations on sexes of the captured individuals
$y_{ijk}^{(1)}$	$y_{ijk}^{(1)} = 1$ if individual i is detected in detector 1 at trap station x_j on occasion k , $y_{ijk}^{(1)} = 0$ if not detected in detector 1
$y_{i\cdot}^{(1)} = \sum_{j=1}^J \sum_{k=1}^K y_{ijk}^{(1)}$	Number of times individual i got detected in detector 1 over J trap stations and K occasions
$y_{ijk}^{(2)}$	$y_{ijk}^{(2)} = 1$ if individual i is detected in detector 2 at trap station x_j on occasion k , $y_{ijk}^{(2)} = 0$ if not detected in detector 2
$y_{i\cdot}^{(2)} = \sum_{j=1}^J \sum_{k=1}^K y_{ijk}^{(2)}$	Number of times individual i got detected in detector 2 over J trap stations and K occasions
n	Number of fully identified individuals, each of them is captured by both the detectors on at least one occasion
$Y_{\text{obs}}^{(1)} = ((y_{ijk}^{(1)}))$	Array of individual-specific capture histories obtained by detector 1 (dimension $n \times J \times K$)
$Y_{\text{obs}}^{(2)} = ((y_{ijk}^{(2)}))$	Array of individual-specific capture histories obtained by & detector 2 (dimension $n \times J \times K$)
$Y^{(1)}$	Zero augmented array of individual-specific capture histories corresponding to detector 1 (dimension $M \times J \times K$)
$Y^{(2)}$	Zero augmented array of individual-specific capture histories corresponding to detector 2 (dimension $M \times J \times K$)
$Y^{(2*)}$	Reordered $Y^{(2)}$ according to L (dimension $M \times J \times K$)
$n_{ij} = \sum_{k=1}^K I(y_{ijk}^{(1)} + y_{ijk}^{(2)} > 0)$	Number of times individual i got detected at trap j on at least one of its sides over K occasions
$n_i = \sum_{j=1}^J n_{ij}$	Number of times individual i got detected on at least one of its sides over J traps and K occasions

Note: Bold symbols represent collections (vectors).

TABLE 3 An example of detection histories for two fully identified individuals and partially identified individuals is presented. The circled 1s indicate the simultaneous captures of an individual by the detectors 1 and 2

	Occasion trap	Detectors 1				Occasion trap	Detectors 2			
		1	2	3	4		1	2	3	4
Fully identified individual 1	1	0	1	0	1	1	0	0	1	0
	2	1	0	0	①	2	0	0	0	①
	3	0	0	1	1	3	1	0	0	0
Fully identified individual 2	1	①	0	0	0	1	①	1	0	0
	2	0	0	0	1	2	0	0	0	0
	3	1	1	0	0	3	0	0	1	0
Partially identified individual	1	1	0	0	1	1	—	—	—	—
	2	0	0	1	0	2	—	—	—	—
	3	0	0	0	0	3	—	—	—	—
Partially identified individual	1	—	—	—	—	1	0	0	1	0
	2	—	—	—	—	2	1	0	0	0
	3	—	—	—	—	3	0	0	1	0

differences between M_2 and M_4 are exactly the same (see Table 4).

We describe the specific models in below.

When detection rates in recorded samples are low due to failure or malfunction of detectors, capture-recapture data may be

comprised of individuals with uncertain identities or "partially identified individuals." Dey et al. (2019) separately accounts for the process of animal arrival within the detection region of a detector and detection process by conditioning on animal arrival—thus modeling

TABLE 4 Specification differences in the four competing models

Model	Trap entry and detection parameter separated?	Sex-specific σ (with sex covariate u)?
M_1	Yes	Yes
M_2	No	Yes
M_3	Yes	No
M_4	No	No

the underlying mechanism by which we obtain different events leading to partial identification.

The probability of animal arrival $\eta_j(\mathbf{s}_i)$ (termed as “trap entry probability”) is modeled as a decreasing function of Euclidean distance $d(\mathbf{s}_i, \mathbf{x}_j) = \|\mathbf{s}_i - \mathbf{x}_j\|$ between individual activity center \mathbf{s}_i and trap station \mathbf{x}_j : $\eta_j(\mathbf{s}_i) = \omega_0 \exp(-d(\mathbf{s}_i, \mathbf{x}_j)^2 / (2\sigma^2))$. Here, ω_0 is regarded as the “baseline trap entry probability” and σ quantifies the rate of decline in trap entry probability as $d(\mathbf{s}_i, \mathbf{x}_j)$ increases. The observation process is parameterized in terms of detection probability ϕ which denotes the probability that any arbitrary individual i is detected by a detector on some occasion k given its arrival at that trap.

The obtained capture history observations $\mathbf{Y}_{\text{obs}}^{(1)} = ((y_{ijk}^{(1)}))_{i,j,k}$ and $\mathbf{Y}_{\text{obs}}^{(2)} = ((y_{ijk}^{(2)}))_{i,j,k}$ from the two detectors 1 and 2 during a spatial capture–recapture survey may not be synchronized as detectors often perform imperfectly. These two observed data arrays are then augmented with “all-zero” capture histories. We denote the zero-augmented data sets by $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$; each of them is of dimension $M \times J \times K$, M being an upper bound of the population size. This also makes the dimension of the likelihood fixed in each iteration of the Markov Chain Monte Carlo algorithm which in turn eases computation. A vector of M latent binary variables $\mathbf{z} = (z_1, \dots, z_M)'$ is introduced where $z_i = 1$ implies that individual i is a member of the population. We assume that each z_i is a Bernoulli random variable with parameter ψ and is independent of other z_i 's. Here ψ is the proportion of individuals that are real and present within \mathcal{Y} . Thus, the true population size N follows the Binomial distribution with parameters M and ψ . The individuals from the two lists obtained from detector 1 and detector 2, respectively, are linked probabilistically by introducing a latent identity variable $\mathbf{L} = (L_1, L_2, \dots, L_M)'$ which is a one-to-one mapping from an index set of individuals captured by detector 2 to $\{1, 2, \dots, M\}$ giving the true index of each of detector 2 individuals. Without loss of generality, the true identity of each individual in the population is defined to be in the row order of the capture histories of detector 1. Then, the rows of detector 2 data set $\mathbf{Y}^{(2)}$ are reordered as indicated by \mathbf{L} to synchronize with the individuals of the detector 1 data set $\mathbf{Y}^{(1)}$. We denote this newly ordered detector 2 data set as $\mathbf{Y}^{(2*)}$.

It is sometimes helpful to introduce a binary covariate on sex category u on spatial animal movement, σ , as in Sollmann et al. (2011). We define σ as a function of the latent structural vector $\mathbf{u} = (u_1, u_2, \dots, u_M)'$: $\sigma(u_i) = \sigma_m$ if $u_i = 1$, that is, individual i is a male; $\sigma(u_i) = \sigma_f$, if $u_i = 0$; that is, individual i is a female. u_i 's are independently and identically distributed Bernoulli random variables with parameter θ , θ being the probability that an arbitrary individual in the population is male. Let $\mathbf{u}_{\text{obs}}(\subset \mathbf{u})$

be a vector of binary observations on sex category of the captured individuals. The vector of latent missing observations in \mathbf{u} is denoted by \mathbf{u}_0 . Assuming that covariate information on individual sex category is available, the joint density of \mathbf{Y}^* and \mathbf{u} under Dey et al. (2019) is the following:

$$f(\mathbf{Y}^*, \mathbf{u}_{\text{obs}} | \theta, \phi, \omega_0, \sigma_m, \sigma_f, \mathbf{u}_0, \mathbf{z}, \mathbf{S}, \mathbf{L}) = \prod_{i=1}^M \left\{ \theta^{u_i} (1-\theta)^{1-u_i} \phi^{y_{i..}} (1-\phi)^{2n_i - y_{i..}} \prod_{j=1}^J \eta_j(\mathbf{s}_i, u_i)^{n_{ij}} \right\}^{z_i} \left\{ (1-\eta_j(\mathbf{s}_i, u_i)) + \eta_j(\mathbf{s}_i, u_i)(1-\phi)^2 \right\}^{K-n_{ij}} \right\} \quad (1)$$

where $\eta_j(\mathbf{s}_i, u_i) = \omega_0 \exp(-d(\mathbf{s}_i, \mathbf{x}_j)^2 / (2\sigma(u_i)^2))$ denotes the probability that an individual i passes through a trap station \mathbf{x}_j on some occasion k , $y_{i..} = y_{i..}^{(1)} + y_{i..}^{(2*)}$, $y_{i..}^{(1)} = \sum_{j=1}^J \sum_{k=1}^K y_{ijk}^{(1)}$, $y_{i..}^{(2*)} = \sum_{j=1}^J \sum_{k=1}^K y_{ijk}^{(2*)}$, $n_{ij} = \sum_{k=1}^K I(y_{ijk}^{(1)} + y_{ijk}^{(2*)} > 0)$ is the number of times individual i got detected on at least one of the detectors over K occasions and $n_i = \sum_{j=1}^J n_{ij}$. The above Equation (1) can

be regarded as zero-inflated Bernoulli density with extra zeros coming from no trap entry. Prior to Dey et al. (2019), Royle (2015) proposed an SCR model for partially identified individuals coming from spatial capture–recapture surveys. The joint density of \mathbf{Y}^* and \mathbf{u} under Royle (2015) is the following:

$$f_{\text{R}}(\mathbf{Y}^*, \mathbf{u}_{\text{obs}} | \theta, p_0, \sigma_m, \sigma_f, \mathbf{u}_0, \mathbf{z}, \mathbf{S}, \mathbf{L}) = \prod_{i=1}^M \left\{ \theta^{u_i} (1-\theta)^{1-u_i} \prod_{j=1}^J p_j(\mathbf{s}_i, u_i)^{y_{ij}} (1-p_j(\mathbf{s}_i, u_i))^{2K-y_{ij}} \right\}^{z_i} \quad (2)$$

where $p_j(\mathbf{s}_i, u_i) = p_0 \exp(-d(\mathbf{s}_i, \mathbf{x}_j)^2 / (2\sigma(u_i)^2))$ denotes the probability that an individual i is detected at \mathbf{x}_j on occasion k . Note that, unlike model (Equation 1), here movement through the detection region is considered inherently as a part of the observation process and p_0 is regarded as “baseline detection probability” and σ , although related to animal movement, is regarded as the *rate of decline in detection probability*. Qualitatively, the absence of ϕ in (Equation 2) distinguishes this model from (Equation 1) and can be regarded as a less general model. Recently, Augustine et al. (2018) extended Royle (2015) by introducing separate parameters to distinguish captures by both detectors and captures by only one of the detectors.

In the absence of the sex covariate \mathbf{u} , the joint density of $\mathbf{Y}^* = (\mathbf{Y}^{(1)}, \mathbf{Y}^{(2*)}) = ((y_{ijk}^{(1)}, y_{ijk}^{(2*)}))$ under Dey et al. (2019) and Royle (2015) is, respectively, as given below:

$$f(\mathbf{Y}^* | \phi, \omega_0, \sigma, \mathbf{z}, \mathbf{S}, \mathbf{L}) = \prod_{i=1}^M \left\{ \phi^{y_{i..}} (1-\phi)^{2n_i - y_{i..}} \prod_{j=1}^J \eta_j(\mathbf{s}_i)^{n_{ij}} \{ (1-\eta_j(\mathbf{s}_i)) + \eta_j(\mathbf{s}_i)(1-\phi)^2 \}^{K-n_{ij}} \right\}^{z_i} \quad (3)$$

$$f_{\text{R}}(\mathbf{Y}^* | p_0, \sigma, \mathbf{z}, \mathbf{S}, \mathbf{L}) = \prod_{i=1}^M \prod_{j=1}^J \left\{ p_j(\mathbf{s}_i)^{y_{ij}} (1-p_j(\mathbf{s}_i))^{2K-y_{ij}} \right\}^{z_i} \quad (4)$$

where $y_{i..} = y_{i..}^{(1)} + y_{i..}^{(2*)}$, $\eta_j(\mathbf{s}_i) = \omega_0 \exp(-d(\mathbf{s}_i, \mathbf{x}_j)^2 / (2\sigma^2))$ denotes the probability that an individual i passes through a trap station \mathbf{x}_j on some occasion k and $p_j(\mathbf{s}_i) = p_0 \exp(-d(\mathbf{s}_i, \mathbf{x}_j)^2 / (2\sigma^2))$ denotes the

probability that an individual i is detected at a trap station \mathbf{x}_i on some occasion k .

The assumed prior distributions for the model parameters are as follows: a uniform distribution over the interval (0,1) for the probability parameters ϕ , ω_0 , p_0 , ψ , and θ ; a uniform distribution over the interval (0, R) for parameters σ , σ_m , and σ_f where R is high enough to expect that it would be impossible for animals to exhibit movement as widely as this scale during sampling. The prior distributions for these model parameters ϕ , ω_0 , p_0 , ψ , and θ , σ , σ_m are assumed to be independent. To ensure that the marginal distribution of the data is well defined, we have assumed proper priors for each of these parameters (Gopalaswamy & Delampady, 2016). We assume a uniform prior over the entire state space \mathcal{V} for each location of the activity center \mathbf{s}_i and that these \mathbf{s}_i 's are independently distributed. The latent variable \mathbf{L} is assumed to have a uniform prior distribution over the permutation space of $\{1, \dots, M\}$. The prior specifications remain the same for all the model fits. The MCMC algorithm used to sample from the respective posterior density under each model is detailed in Appendix S1D. Thus, we have the four models,

M_1 : Model with density (Equation 1),

M_2 : Model with density (Equation 2),

M_3 : Model with density (Equation 3), and

M_4 : Model with density (Equation 4).

Model selection tools are used to find the “best” fitted model among a set of competing models. To provide an understanding of what such a best model is, Shmueli (2010) discusses two broad modelling philosophies—explanatory and predictive. Explanatory modeling of the data set is only relevant to obtain the most accurate representation of the underlying theory, whereas predictive modelling seeks to minimize the combination of bias and estimation variance. Hence, a model that explains the data well may not have the best predictive ability and vice versa. This naturally leads to two performance aspects by which a model can be assessed—(a) how well does the model explain the observed data set, that is, to what degree of confidence can it be thought of as a model from which the data generated, (b) how good is its predictive ability.

On such criteria, a wide range of model selection tools have thus emerged in the applied statistical literature (see Gelman, Hwang, & Vehtari, 2014; Höge, Guthke, & Nowak, 2019; Höge, Wöhling, & Nowak, 2018; Hooten & Hobbs, 2015; Kass & Raftery, 1995; for examples). In all these cases, however, we are assuming that the “true” model is in the list of competing models. Such classes of problems are formally categorized as ‘ \mathcal{M} -closed’ (Bernardo & Smith, 1994; Clarke, Clarke, & Yu, 2013; Vehtari & Ojanen, 2012) and is generally considered to be well-studied (Clarke, Clarke, & Yu, 2013).

However, for the practitioner, the applicability of such guidelines for model selection is not clear because often such guidelines are formed using specific examples (Hooten & Hobbs, 2015) or based on asymptotic arguments in a theoretic sense (for example, Shibata, 1989; Stone, 1977; Watanabe, 2010). Even when the purpose of model selection is clear (explanatory or predictive) and also

the model selection criterion is fixed, what techniques practically work is not clear. Practitioners often wish to confront fairly complex models (e.g., hierarchical models) with their data, and it is often a challenge to directly apply a candidate set of model selection tools without incorporating some approximations. Given these approximations due to model complexity and with limited sample sizes, the validity of such general guidelines (Hooten & Hobbs, 2015) remains to be tested in practice. Therefore, in this study, we apply a range of model selection tools and test the validity of such general model selection guidelines in the context of a popular class of models, spatial capture–recapture models, used for estimating animal density of some of the world's most iconic species (Broekhuis & Gopalaswamy, 2016; Elliot & Gopalaswamy, 2017; Royle, Karanth, Gopalaswamy, & Kumar, 2009; Sollmann et al., 2011).

2.2 | Candidate model selection tools

We have considered four different Bayesian model selection methods for application and evaluation: *Bayes factors*, *Deviance Information Criterion* (DIC), *Watanabe-Akaike information criterion* (WAIC), and *posterior predictive loss*. Two popular model selection tools (AIC and BIC; Burnham & Anderson, 2002) are not used here because they impose restrictive assumptions on the parameter space as the sample size increases—situations often encountered in many hierarchical models (Royle & Dorazio, 2008). For example, in the SCR models we study here, the concept of “number of parameters” is unclear and we therefore cannot apply criteria such as the AIC and the BIC directly.

We clarify our aim here is to discriminate between hypotheses (models) and not to conduct hypothesis tests of the model parameters. In the hypothesis testing paradigm that applies in our setup, it usually requires for models to be nested so that statistical tests (such as likelihood ratio tests) between different parameter values can be performed. However, this limitation does not apply to model selection under the hypothesis discrimination paradigm.

2.2.1 | Bayes factors

Model comparison using Bayes factors requires the computation of the marginal likelihood, which involves the integration $m(\mathbf{Y}|M_i) = \int f(\mathbf{Y}|\boldsymbol{\mu}, M_i) \pi(\boldsymbol{\mu}) d\boldsymbol{\mu}$ where $f(\mathbf{Y}|\boldsymbol{\mu}, M_i)$ denotes the model likelihood and $\pi(\boldsymbol{\mu})$ denotes the prior density of the parameters $\boldsymbol{\mu}$ under M_i . This integration is difficult to compute in practice unless the models are very simple in structure, which is often not the case in ecology. Therefore, computation of marginal likelihood continues to be an active area of research in Statistics (Wang & Meng, 2016; Wang et al., 2018).

Estimation of marginal likelihood of data

Under our model settings, $\mathbf{Y} = (\mathbf{Y}^*, \mathbf{u}_{\text{obs}})$ for models M_1 , M_2 , and $\mathbf{Y} = \mathbf{Y}^*$ for the other models M_3 , M_4 . $\boldsymbol{\mu}$ denotes the collection of all parameters and latent variables for each model as a generic notation. Specifically, let $\boldsymbol{\mu} = (\boldsymbol{\mu}_p, \boldsymbol{\mu}_s)$, where $\boldsymbol{\mu}_p$ is the collection of scalar parameters and $\boldsymbol{\mu}_s$ is the collection of all latent variables. The Gelfand-Dey estimator of marginal likelihood of data $m(\mathbf{Y})$ is expressed as:

$$\hat{m}_{\text{GD}}(\mathbf{Y}) = \left[\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \frac{g(\boldsymbol{\mu}^{(d)})}{f(\mathbf{Y}|\boldsymbol{\mu}^{(d)})\pi(\boldsymbol{\mu}^{(d)})} \right]^{-1}, \quad (5)$$

where $\{\boldsymbol{\mu}^{(d)}: d = 1, \dots, N_{\text{iter}}\}$ is a set of MCMC draws from the posterior $\pi(\boldsymbol{\mu}|\mathbf{Y})$ and $g(\boldsymbol{\mu})$ is a tuning density. It is to be noted that by specifying $g(\boldsymbol{\mu}) = \pi(\boldsymbol{\mu})$ in (Equation 5), we obtain the harmonic mean estimator of the marginal likelihood

$$\hat{m}_{\text{HM}}(\mathbf{Y}) = \left[\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \frac{1}{f(\mathbf{Y}|\boldsymbol{\mu}^{(d)})} \right]^{-1}. \quad (6)$$

Details on these estimators and their properties can be found in Gelfand and Dey (1994), Kass and Raftery (1995). The harmonic mean estimator is easy to compute by just calculating the model likelihood at each MCMC sample draws from the posterior distribution. Although it is known that the harmonic mean estimator is a consistent estimator of the marginal likelihood, as noted in Kass and Raftery (1995) and Newton and Raftery (1994), it can have a simulation pseudo-bias (Lenk, 2009). This bias occurs due to the limitations in numerical computations. Indeed, our study here aims to assess how and whether these potential biases in computation influence the model selection process for the particular class of models we consider—partial identity SCR models, in our case.

For our problem, the computation of (Equation 5) requires us to obtain the integrated likelihoods (marginals) under the different models that we consider. This becomes particularly tricky in the presence of high-dimensional latent variables such as \mathbf{u}_0 , \mathbf{z} , \mathbf{S} , and \mathbf{L} , which are elements of $\boldsymbol{\mu}_s$.

We have developed a novel approach to compute an estimate of the marginal likelihood. Here we partition the parameters of a model into two sets (real-valued scalar and high-dimensional structural parameters, which are denoted by $\boldsymbol{\mu}_p$ and $\boldsymbol{\mu}_s$, respectively) and then propose two approximating strategies to account for the high-dimensional structural parameters, $\hat{\boldsymbol{\mu}}_s$, before applying the Gelfand-Dey estimator. This general approach of partitioning can be applied to many other classes of hierarchical models as well and not limited to SCR models.

We describe below two *approximating* approaches to compute the Gelfand-Dey estimator: the maximum a posteriori (MAP) approximation approach and the integrated likelihood (IL) approach.

Approach 1: MAP approximation

In this approach, we fix the high-dimensional variables at their MAP estimates $\hat{\boldsymbol{\mu}}_s$, assuming that their posterior distributions are well summarized by these estimates, which are derived from the MCMC draws. The Gelfand-Dey estimator is then computed using the formula,

$$\hat{m}_{\text{GD}}(\mathbf{Y}) = \left[\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \frac{g(\boldsymbol{\mu}_p^{(d)})}{f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d)}, \hat{\boldsymbol{\mu}}_s)\pi(\boldsymbol{\mu}_p^{(d)})} \right]^{-1}. \quad (7)$$

where $\{(\boldsymbol{\mu}_p^{(d)}, \boldsymbol{\mu}_s^{(d)}): d = 1, \dots, N_{\text{iter}}\}$ is a set of MCMC draws from the posterior $\pi(\boldsymbol{\mu}_p, \boldsymbol{\mu}_s|\mathbf{Y})$. We begin with $(\boldsymbol{\mu}_p^{(d_0)}, \boldsymbol{\mu}_s^{(d_0)})$ as an initial estimate of $(\boldsymbol{\mu}_p, \boldsymbol{\mu}_s)$ where $f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d_0)}, \boldsymbol{\mu}_s^{(d_0)})\pi(\boldsymbol{\mu}_p^{(d_0)}, \boldsymbol{\mu}_s^{(d_0)}) = \max_d \{f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d)}, \boldsymbol{\mu}_s^{(d)})\pi(\boldsymbol{\mu}_p^{(d)}, \boldsymbol{\mu}_s^{(d)})\}$.

This estimate of the posterior mode of $(\boldsymbol{\mu}_p, \boldsymbol{\mu}_s)$ may not be optimal since in our high-dimensional parameter setting, an MCMC sample of a practical size may not be enough to extensively explore the posterior surface. We, therefore, fix one of the parameters $\boldsymbol{\mu}_s = \boldsymbol{\mu}_s^{(d_0)}$ and explore the posterior surface to find d_1 such that $f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d_1)}, \boldsymbol{\mu}_s^{(d_0)})\pi(\boldsymbol{\mu}_p^{(d_1)}, \boldsymbol{\mu}_s^{(d_0)}) = \max_d \{f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d)}, \boldsymbol{\mu}_s^{(d_0)})\pi(\boldsymbol{\mu}_p^{(d)}, \boldsymbol{\mu}_s^{(d_0)})\}$. In this way, we obtain an improved MAP estimate of $(\boldsymbol{\mu}_p, \boldsymbol{\mu}_s)$, $(\boldsymbol{\mu}_p^{(d_1)}, \boldsymbol{\mu}_s^{(d_0)})$, if $f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d_1)}, \boldsymbol{\mu}_s^{(d_0)})\pi(\boldsymbol{\mu}_p^{(d_1)}, \boldsymbol{\mu}_s^{(d_0)}) > f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d_0)}, \boldsymbol{\mu}_s^{(d_0)})\pi(\boldsymbol{\mu}_p^{(d_0)}, \boldsymbol{\mu}_s^{(d_0)})$.

Similarly, we then fix $\boldsymbol{\mu}_p = \boldsymbol{\mu}_p^{(d_1)}$ and find $\boldsymbol{\mu}_s^{(d_2)}$. This procedure is continued iteratively to eventually give us the best MAP estimate of the posterior mode $(\hat{\boldsymbol{\mu}}_p, \hat{\boldsymbol{\mu}}_s)$. Suitable transformations of the parameters ensure that all the points in $\{(\boldsymbol{\mu}_p^{(a)}, \boldsymbol{\mu}_s^{(b)}): a, b = 1, \dots, N_{\text{iter}}; a \neq b\}$ belong to the posterior support. A single MCMC chain is to be used for the algorithm, and we have shown that the obtained MAP estimate using our algorithm will be better than the usual estimate $(\boldsymbol{\mu}_p, \boldsymbol{\mu}_s)$ (see Appendix S1A). In theory, we can also increase sample size by merging all the chain outputs after convergence.

Approach 2: Integrated likelihood (IL) approximation

Ideally, we would like to compute the marginal likelihood $m(\mathbf{Y})$ by integrating out all the latent variables with respect to their corresponding prior distributions from the model likelihoods. However, in the case of model likelihoods (Equations 1–4), this integration is computationally difficult for the permutation vector \mathbf{L} because of its very large support. The integration over the variables \mathbf{u}_0 and \mathbf{z} can be performed analytically. The integration over \mathbf{S} is evaluated numerically by partitioning the region \mathcal{V} into a sufficiently fine grid and then evaluating a Riemann sum (as direct integration cannot be expressed in a closed form). This integrated likelihood can then be used in (Equation 5) for estimating the marginal likelihood

$$\hat{m}_{\text{GD}}(\mathbf{Y}) = \left[\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \frac{g(\boldsymbol{\mu}_p^{(d)}, \mathbf{L}^{(d)})}{f(\mathbf{Y}|\boldsymbol{\mu}_p^{(d)}, \mathbf{L}^{(d)})\pi(\boldsymbol{\mu}_p^{(d)}, \mathbf{L}^{(d)})} \right]^{-1}. \quad (8)$$

One downside of the IL approximation approach is the lack of clarity about the interdependencies between the latent variables \mathbf{u}_0 , \mathbf{z} , \mathbf{S} , and \mathbf{L} after carrying out the integrations. The derivations of the integrated likelihoods for each of the four models M_1 – M_4 are given in Appendix S1B by ignoring any possible interdependencies. We have also assessed the robustness of the Gelfand-Dey estimator by computing the marginal likelihood estimates using different tuning densities, for example, multivariate normal density, multivariate-t density with varying degrees of freedom, and the truncated normal density following the suggestion of Geweke (1999). These technical details are described in Appendix S1C.

2.2.2 | Deviance information criterion

Deviance is defined as $D(\boldsymbol{\mu}) = -2\log f(\mathbf{Y}|\boldsymbol{\mu})$. The deviance information criterion (DIC) is then defined as $\text{DIC} = D(\hat{\boldsymbol{\mu}}) + 2p_{\text{DIC}}$, where $\hat{\boldsymbol{\mu}}$ is an estimate of $\boldsymbol{\mu}$. The term p_{DIC} can be viewed as effective number of parameters, which is a bias correction in evaluating a model's predictive

accuracy. Two versions of p_{DIC} are generally used (Gelman, Carlin, et al., 2014; Spiegelhalter, Best, Carlin, & van der Linde, 2002):

$$p_{\text{DIC1}} = \mathbb{E}_{\mu|Y}[D(\mu)] - D(\hat{\mu}) = 2(\log f(\mathbf{y}|\hat{\mu}) - \mathbb{E}_{\mu|Y}[\log f(\mathbf{y}|\mu)]), \quad (9)$$

$$p_{\text{DIC2}} = 2\text{Var}_{\mu|Y}[\log f(\mathbf{y}|\mu)].$$

Although p_{DIC1} has been considered to be numerically more stable, p_{DIC2} has the advantage of being always positive (Gelman, Hwang, & Vehtari, 2014). A model with smaller DIC value is preferred. Model comparison using DIC is not invariant to parameterization and depends on the components of the model likelihood to be considered as the likelihood. Spiegelhalter et al. (2002) suggests that practitioners carefully decide on the parameters of interest so that they can avoid this potential pitfall. This advice is often difficult to implement in practice, especially when there exists inherent ambiguity in the interpretation of latent parameters. We note that Celeux, Forbes, Robert, and Titterton (2006) suggests several forms for the DIC that can be used for different hierarchical models but does not recommend any particular form as the best.

For the computation of deviance, we have use the MAP estimate of μ to obtain $\hat{\mu}$ instead of the posterior mean, due to the presence of binary latent variables and unknown permutation vectors in the likelihood. We have then computed two versions of p_{DIC} (Gelman, Carlin, et al., 2014; Hooten & Hobbs, 2015) using the MCMC draws $\{\mu^{(d)}: d = 1, \dots, N_{\text{iter}}\}$ from $\pi(\mu|Y)$ as follows:

$$\hat{p}_{\text{DIC1}} = 2 \log f(\mathbf{Y}|\hat{\mu}) - \frac{2}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \log f(\mathbf{Y}|\mu^{(d)}),$$

$$\hat{p}_{\text{DIC2}} = 2 \left[\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \left(\log f(\mathbf{Y}|\mu^{(d)}) - \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \log f(\mathbf{Y}|\mu^{(d)}) \right)^2 \right] \quad (10)$$

2.2.3 | Watanabe-Akaike information criterion

The Watanabe-Akaike information criterion (WAIC) is a Bayesian version of the AIC as it uses the posterior predictive distribution of the data to estimate the out-of-sample predictive accuracy of the model. Watanabe (2010) introduced the WAIC criterion based on the assumption of independence between the data points and has shown its asymptotic equivalence with cross-validation. In our model formulations, we have assumed that the different data points correspond to capture-recapture data set for each of the M individuals.

The WAIC is then defined as $\text{WAIC} = -2 \sum_{i=1}^M \log \mathbb{E}_{\mu|Y}(f(\mathbf{Y}_i|\mu)) + 2p_{\text{WAIC}}$. A model with smaller WAIC value is preferred. In computing WAIC, we partition data Y in terms of individuals (Y_1, Y_2, \dots, Y_M) . We compute the two commonly used versions of p_{WAIC} (Hooten & Hobbs, 2015) using MCMC draws $\{\mu^{(d)}: d = 1, \dots, N_{\text{iter}}\}$ from $\pi(\mu|Y)$ as follows:

$$\hat{p}_{\text{WAIC1}} = 2 \sum_{i=1}^M \left\{ \log \left(\frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} f(\mathbf{Y}_i|\mu^{(d)}) \right) - \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \log f(\mathbf{Y}_i|\mu^{(d)}) \right\},$$

$$\hat{p}_{\text{WAIC2}} = \sum_{i=1}^M \left\{ \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \left(\log f(\mathbf{Y}_i|\mu^{(d)}) - \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \log f(\mathbf{Y}_i|\mu^{(d)}) \right)^2 \right\}. \quad (11)$$

We propose another version for p_{WAIC} based on absolute error loss, recognizing that large variability in $\mu^{(d)}$ and the magnitude of squared errors themselves may have an impact on the efficiency of p_{WAIC} of the above forms:

$$\hat{p}_{\text{WAIC3}} = 2 \sum_{i=1}^M \left\{ \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \left| \log f(\mathbf{Y}_i|\mu^{(d)}) - \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \log f(\mathbf{Y}_i|\mu^{(d)}) \right| \right\}. \quad (12)$$

From a decision-theoretic perspective, we think it is interesting to compare WAIC2, which is based on a square error loss function with WAIC3 that is based on an absolute error loss function. The comparison based on a simulated data analysis is described later in Section 4.

2.2.4 | Posterior predictive loss

Gelfand and Ghosh (1998) derived a model selection criterion, popularly known as the posterior predictive loss criterion, by adopting a decision-theoretic approach for measuring predictive accuracy of a model. The posterior predictive loss D_w criterion (based on a square error loss function) is composed of squared bias, variance of a new prediction, and a parameter w . The parameter w indicates the relative weight given to the loss for departure of the new predicted value from the observed data against departure of the same from the new data. In practice, this weight is taken to be very large, and asymptotically as $w \rightarrow \infty$, the D_∞ criterion is obtained as follows under our model setting:

$$D_\infty = \sum_{i=1}^{2MJK} (y_{i,\text{vec}} - \mathbb{E}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}}))^2 + \sum_{i=1}^{2MJK} \text{Var}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}}), \quad (13)$$

where $\mathbf{Y}_{\text{vec}} = (y_{1,\text{vec}}, y_{2,\text{vec}}, \dots, y_{2MJK,\text{vec}})'$ is a vector of length $2MJK$ obtained by vectorizing observed data array \mathbf{Y} and $\mathbf{Y}_{\text{rep}} = (y_{1,\text{rep}}, y_{2,\text{rep}}, \dots, y_{2MJK,\text{rep}})'$ represents the vector of all these replicates. The first term in the D_∞ criterion (see Equation 13) is the goodness-of-fit term, while the second term can be interpreted as a penalty term for model complexity. The model with the smallest D_∞ is to be preferred. In our analysis, the two data sets are obtained by vectorizing the two binary data arrays.

We compute the above expectation $\mathbb{E}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}})$ and variance $\text{Var}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}})$ using the MCMC draws. Given an MCMC sample $\mu^{(d)}$: $d = 1, \dots, N_{\text{iter}}$ from $\pi(\mu|Y)$, we simulate $\mathbf{Y}_{\text{rep}}^{(d)}$ from $f(\mathbf{Y}|\mu^{(d)})$ for each $d = 1, \dots, N_{\text{iter}}$. For instance, in model M_1 , μ denotes the collection of the parameters $\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_\theta, \mathbf{u}_0, \mathbf{z}, \mathbf{S}, \mathbf{L}$. Then $\mathbb{E}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}}) \approx N_{\text{iter}}^{-1} \sum_{d=1}^{N_{\text{iter}}} y_{i,\text{rep}}^{(d)}$

and $\text{Var}(y_{i,\text{rep}}|\mathbf{Y}_{\text{vec}}) \approx N_{\text{iter}}^{-1} \sum_{d=1}^{N_{\text{iter}}} \left(y_{i,\text{rep}}^{(d)} - N_{\text{iter}}^{-1} \sum_{d=1}^{N_{\text{iter}}} y_{i,\text{rep}}^{(d)} \right)^2$. We summarize

the various model selection methods and their variants in Table 5. Considering all these model selection tools and their variants (from approximation approaches to setting tuning densities), our evaluation is carried out on 25 unique tools. By no means is this an exhaustive list of candidate tools because there are a large number of tools

TABLE 5 Bayesian model selection methods used in this study

Sl. no.	Model selection method	Variant	Approximation method	Choices of tuning density	Eq. No.
1.	Bayes factor	Gelfand-Dey estimator	MAP	Multivariate normal density, multivariate-t density with degrees of freedom 10, 100, 500, 1,000, 10,000 and truncated multivariate normal density with confidence coefficients 0.90, 0.95, 0.99	(7)
2.	Bayes factor	Gelfand-Dey estimator	IL	-Do-	(8)
3.	Bayes factor	Harmonic mean estimator	—	—	(6)
4.	DIC	p_{DIC1}	MAP	—	(10)
5.	DIC	p_{DIC2}	MAP	—	(10)
6.	WAIC	p_{WAIC1}	—	—	(11)
7.	WAIC	p_{WAIC2}	—	—	(11)
8.	WAIC	p_{WAIC3}	—	—	(12)
9.	Posterior predictive loss	—	—	—	(13)

available. Our purpose here really is to assess some popular tools and assess their efficacies on a particular class of models.

3 | EVALUATION OF THE PERFORMANCE OF MODEL SELECTION METHODS

3.1 | Simulation design and simulation scenarios

We have conducted simulations for 12 scenarios (provided in Table 6) grouped into 2 equal sized sets, to assess the performance of the models proposed here. We set $\sigma_m = 0.3$ and $\sigma_f = 0.15$ for the first set of 6 scenarios, $\sigma_m = 0.4$ and $\sigma_f = 0.2$ for the second set of 6 scenarios. We set $(\omega_0, \phi) = \{(0.01, 0.3), (0.05, 0.3), (0.05, 0.5), (0.03, 0.8), (0.01, 0.9), (0.05, 0.9)\}$, which gives us 6 different scenarios for

each of the two sets corresponding to the values taken by ω_0 and ϕ . We assume that a total of 100 individuals are residing inside the state space of which 40 are male. Each of the simulation experiments is conducted within a rectangular state space of dimension 5 unit \times 7 unit (Figure 1), after setting a buffer of 1 unit in both horizontal and vertical directions, a 10×16 trapping array of total $J = 160$ trap stations has been set (trap spacing is 0.3 unit on X axis and 0.3125 unit on Y axis). This meets the requirement suggested in Karanth and Nichols (2017). Each of the traps remains active for $K = 50$ sampling occasions simultaneously. For parameter estimation, we set the

TABLE 6 Parameter specifications corresponding to different simulation scenarios

Scenario	M	N	N_{Male}	ω_0	ϕ	σ_m	σ_f
1	400	100	40	0.01	0.3	0.3	0.15
2	400	100	40	0.01	0.9	0.3	0.15
3	400	100	40	0.01	0.3	0.4	0.20
4	400	100	40	0.01	0.9	0.4	0.20
5	400	100	40	0.03	0.8	0.3	0.15
6	400	100	40	0.03	0.8	0.4	0.20
7	400	100	40	0.05	0.3	0.3	0.15
8	400	100	40	0.05	0.5	0.3	0.15
9	400	100	40	0.05	0.9	0.3	0.15
10	400	100	40	0.05	0.3	0.4	0.20
11	400	100	40	0.05	0.5	0.4	0.20
12	400	100	40	0.05	0.9	0.4	0.20

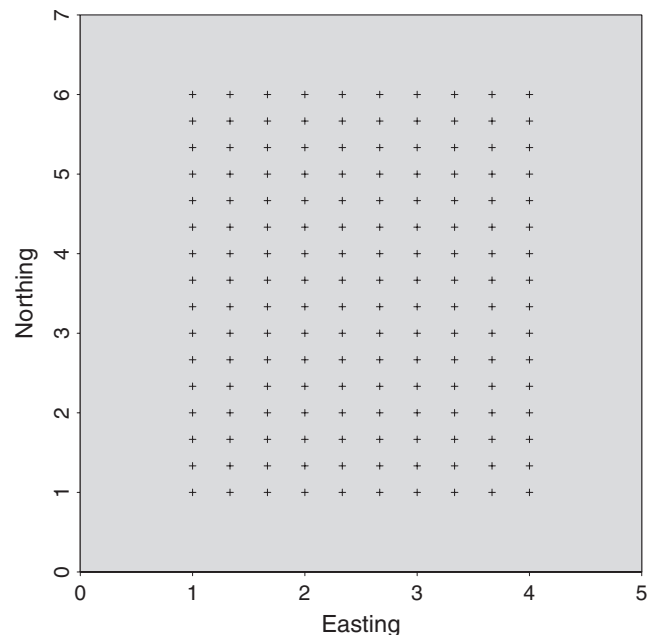


FIGURE 1 Array of trap locations (denoted by “+”) within the state space (0.5) \times (0.7)

maximum possible number of individuals present in the population (M) at 400 for all the scenarios. Since MCMC approaches are computationally time consuming, we set the data-generating parameters to generate data sets of varying levels of information content. This strategy ensured that we capture trends in model selection performance by the different methods. The experiment is repeated $n_{\text{sim}} = 10$ times. The MCMC samples for each of the parameters are obtained (each of length 30,000), and the estimates are computed using those chains after a burn-in of 10,000. It takes approximately 3 days to run a single chain of 30,000 iterations using our R code for model M_1 , approximately 2 days for model M_2 and M_3 , and approximately 1 day for model M_4 using our R code on a DELL Precision Rack 7,910 Server with 40 cores and 512 GB RAM at a clock speed of 2.20 GHz.

Capture–recapture data sets are simulated independently under each of the 12 simulation scenarios (Table 6) under model M_1 . Recall that, model M_1 corresponds to the statistical model in (Equation 1) with σ parameter modeled in terms of individual covariate on sex category \mathbf{u} (see Section 2.1). Then, each simulated data set is fitted with all the four competing model M_1, M_2, M_3 , and M_4 . For a better mixing during the MCMC run, we transform the components of $\boldsymbol{\mu}_p$ so that the transformed parameter space is the entire Euclidean space (details are given in Appendix S1C).

3.2 | Defining performance measures

3.2.1 | Probability of selecting the true model

In our study, since all the data sets are simulated from model M_1 , it is considered as the true model. We have computed the proportion of times a model selection method chooses M_1 as the best out of $n = 10$ simulations. This proportion will serve as an estimate for the probability of selecting the true model. Additionally, we have also computed the proportion of times a model selection method chooses M_2, M_3 , and M_4 , respectively. The computed proportions of selecting these models from the simulation study will indicate the efficacy of the model selection methods.

3.2.2 | Assessing the quality of parameter estimation

The abundance parameter N carries a lot of significance in ecology and conservation. Due to its importance, ecologists place their interests in the robustness and accuracy of its estimate, and will therefore rely on a model selection method that will achieve this by selecting that model from the model set that will provide the most reliable estimate of N . The precision and accuracy of the parameter estimate indicates the quality of the model fit, and we assess this by computing the *average root mean square error* (average RMSE).

Suppose $\{\mu^{(td)}; d = 1, \dots, N_{\text{iter}}\}$ denotes a set of MCMC draws from the posterior distribution of an arbitrary parameter μ for the t -th simulated data set, $t = 1, \dots, n_{\text{sim}}$. Mean square error (MSE) of μ for the

t -th simulated data set is estimated as $\text{MSE}(\mu, t) = N_{\text{iter}}^{-1} \sum_{d=1}^{N_{\text{iter}}} (\mu^{(td)} - \mu)^2$.

Average RMSE is calculated by first averaging the estimated MSEs of different simulations and then taking the square root of the average:

$$\text{Average RMSE}(\mu) = \sqrt{n_{\text{sim}}^{-1} \sum_{t=1}^{n_{\text{sim}}} \text{MSE}(\mu, t)}.$$

Let us now summarize our experiment. The quantities related to the various Bayesian model selection methods (Section 2.2) are computed for each of the four models for the simulated data and analysis sets. We compare the conclusions drawn from the computed proportions with the conclusions drawn from average RMSEs of the parameters to study the behavior of the model selection methods with respect to varying information content. We also generate pairwise correlation plots from the MCMC draws to study the extent of parameter redundancy between various pairs of parameters as a consequence of lack of information content in the data.

4 | RESULTS OF THE SIMULATION STUDY

Figure 2 shows the proportion of times different estimators of Bayes factor favors any particular model. The subplots have different scenarios 1–12 in the x-axis and the competing models M_1 – M_4 in the y-axis. Plot (a) and Plot (b) correspond to the Gelfand–Dey estimator of the Bayes factor favoring any particular model using (i) the MAP approximation approach and a multivariate normal density for tuning density g , (ii) the integrated likelihood approximation approach and a multivariate normal density for tuning density g , respectively. Plot (c) corresponds to the harmonic mean estimator of the Bayes factor. Plots (d–i) correspond to WAIC1, WAIC2, WAIC3, DIC1, DIC2, and posterior predictive loss, respectively.

Our results suggest that the choice of the tuning density in computation of Gelfand–Dey estimator had no impact on model selection (Figure 2). Thus, we focus only on the results corresponding to the normal tuning density. We observe that the Bayes factor with the GD-MAP approximation favors M_4 in 70% of the times under scenario 1 and favors M_3 in at least 70% of the times under scenarios 2, 3, 4, and 7. Bayes factor (GD-MAP) is in favor of M_1 more than 80% of the times under all the remaining scenarios 5–6 and 8–12. As seen here, the GD-IL approximation performs slightly worse than the GD-MAP approximation (Figure 2). However, surprisingly, we observe that the model choices by the harmonic mean estimator of Bayes factor performs well and favors the true model M_1 in majority of the scenarios.

We have considered three forms of the WAIC and two forms of the DIC. The corresponding plots are shown in Figure 2(d–h). WAIC1, WAIC3, DIC1, and DIC2 exhibit very similar tendencies in their model choices by favoring M_3 under all the scenarios except scenarios 1 and 3. These four methods favor M_4 under scenario 1.

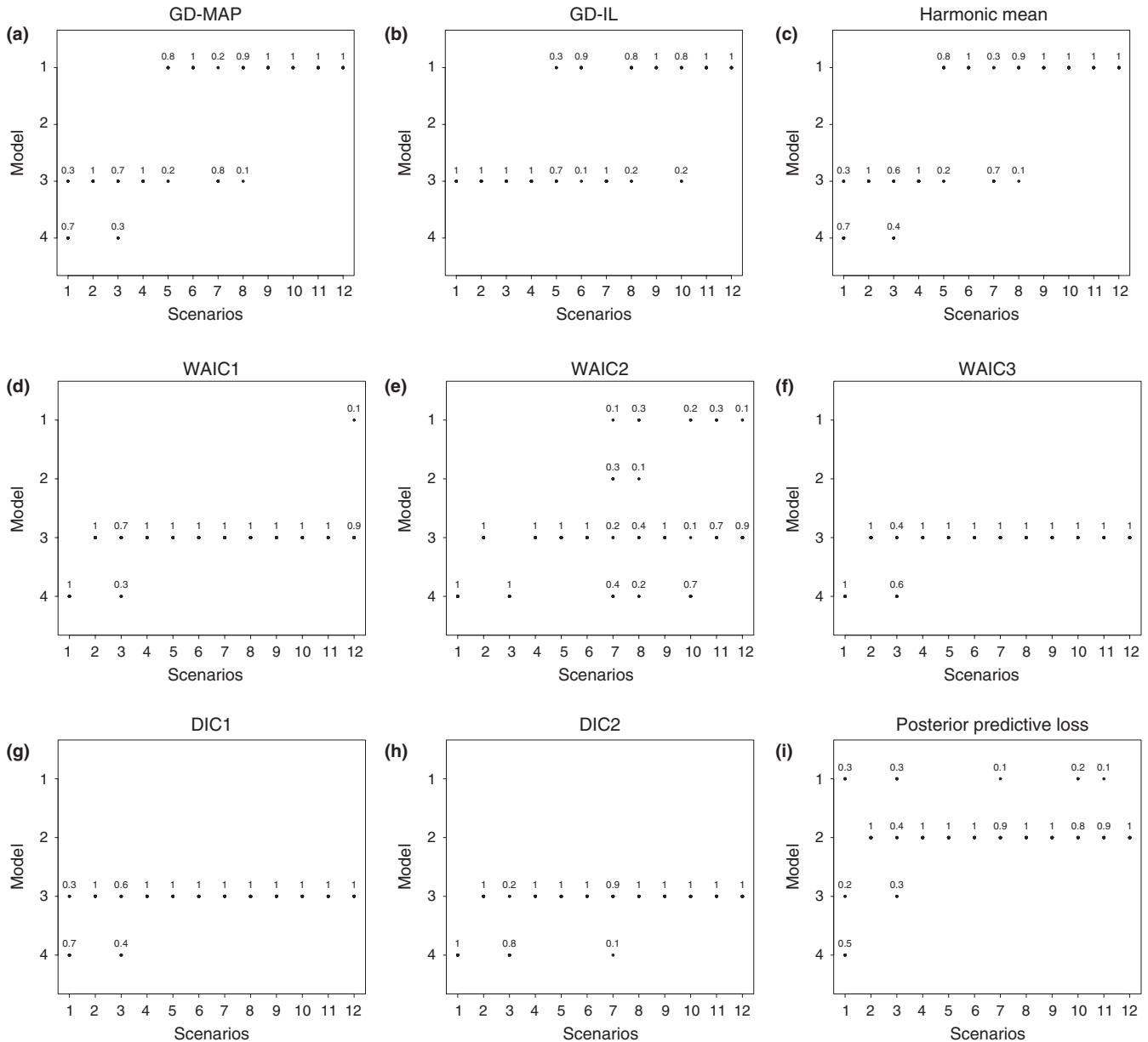


FIGURE 2 Plot (a): The proportion of times Gelfand-Dey estimator of Bayes factor favors any particular model using the MAP approximation approach and a multivariate normal density for g . Plot (b): The proportion of times Gelfand-Dey estimator of Bayes factor favors any particular model using the integrated likelihood approximation approach and a multivariate normal density for g . Plot (c): The proportion of times harmonic mean estimator of Bayes factor using the favors any particular model. Plots (d)–(i) correspond to WAIC1, WAIC2, WAIC3, DIC1, DIC2, and posterior predictive loss, respectively

Under scenario 3, WAIC1 and DIC1 favor M_3 in majority of the times, whereas WAIC3 and DIC2 favor M_4 . Like Bayes factor, these model selection methods (DICs and WAICs) also tend to favor simpler models under scenarios 1 and 3. The WAIC2 largely agrees with the other WAICs (and DICs) but more often selects the true model when data sets are more informative (Figure 2d–f). In general, DICs and WAICs seem to discourage the presence of high-dimensional latent variables.

The plot showing the proportion of different model choices by posterior predictive loss is given in Figure 2i. The posterior predictive loss criterion D_∞ favors models with individual sex covariates

(M_1 and M_2). The posterior predictive loss also appears to select the true model some of the times, even when there is not sufficient information in the data (scenarios 1 and 3).

In Figure 3, we see that the average RMSE of N is substantially higher in scenarios 1–3 relative to scenarios 4–12. We also have observed a substantial amount of correlation between some pairs of parameters involving N in the MCMC chains under scenarios 1–3 ($r(N, \theta) \approx -0.7$, $r(N, \sigma_f) \approx -0.42$). However, these correlations decrease for the other scenarios, likely due to increased information content in the data. For interested readers, we have also provided additional plots for the model selection methods (Figures 1–6), the RMSE plots

for all the parameters (Figures 7 and 8), and a representative set of scatter plots (Figures 9–14) in the Appendix S1E.

5 | DISCUSSION AND CONCLUSIONS

Contemporary practice of ecology and conservation biology relies largely on the use of model selection for hypotheses discrimination (Johnson & Omland, 2004). Simultaneously, there has been major growth in the use of hierarchical models in ecology, especially within the realm of Bayesian inference (Kéry & Royle, 2015; Royle & Dorazio, 2008). These models have now enabled statistical ecologists to fairly easily formulate complex ecological models, elegantly deal with the sampling process and also fit these complex models using powerful tools such as the MCMC (Kéry & Royle, 2015). However, the lack of availability of ready-made model selection tools when practicing the Bayesian inference has sometimes motivated ecologists to continue using the likelihood-based inferences, merely because one can use well-known model selection tools such as the AIC (Burnham & Anderson, 2002) for inference.

To provide a context for this argument, in the spatial capture–recapture literature, we have essentially seen the development of three important likelihood functions: (1) Borchers and Efford (2008), (2) Royle et al. (2009) and (3) Royle, Sutherland, Fuller, and Sun (2015). Inferences for the models (1) and (3) are by maximizing the likelihood, while the inference for (2) is Bayesian. We note with interest that one of the reasons for the development of (3) was motivated on the pretext that model selection is much easier (using known tools such as the AIC) for practitioners using the likelihood approach, in spite of the problem having been already solved in the Bayesian context (Royle et al., 2013). It is specially of concern in the context of the models we study here in that investigators may be forced to integrate out \mathbf{S} (activity centers of individuals) in order to construct tractable likelihoods and thus oversimplifying ecological reality. For example, by retaining the activity centers it is possible for investigators to confront open SCR models (Gardner, Sollmann, Kumar, Jathanna, & Karanth, 2018). We therefore hope that our study motivates the continued use of Bayesian methods by investigators for all its advantages, instead of opting out of them merely for the sake of using simple model selection methods.

In this study, we have tried to implement some selected Bayesian model selection methods on a specific class of advanced Bayesian SCR models (Royle, 2015—with and without sex covariates; Dey et al. (2019)—with and without sex covariates) dealing with partially identified individuals. We have found our Bayes factor implementation using the Gelfand–Dey estimator (using the MAP approximation approach) to be the preferred choice as a model selection method over a wide range of simulation scenarios. This approach appears to work particularly well when information content in the data is moderate to high. The IL approximation approach also worked well but not as well as the MAP approximation approach perhaps because

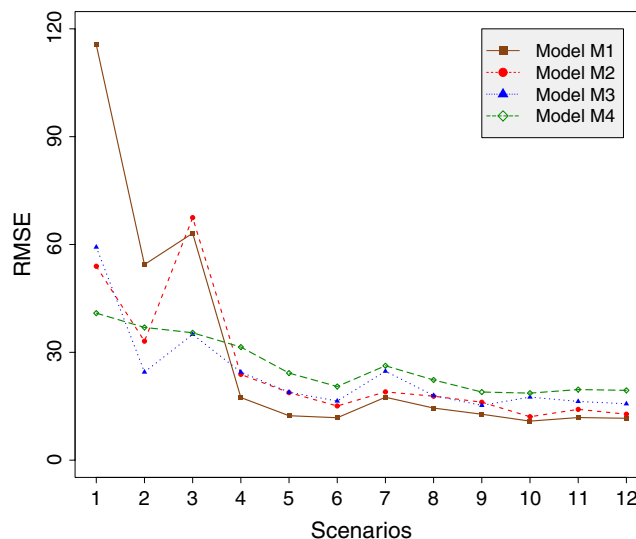


FIGURE 3 Plot of average RMSE estimates of N over different simulation scenarios

there exists interdependency between some of the latent variables which are not accurately captured during marginalization.

However, implementing Bayes factors for model selection using the Gelfand–Dey estimator (with MAP approximation) is not very straightforward and requires investigators to incorporate some code during model fitting by MCMC. Interestingly, our study demonstrates that obtaining Bayes factors using the harmonic mean approach for marginal likelihood computation is less demanding but yet serves as a very good model selection method. This finding deviates from the popular view among applied scientists that it is futile to estimate the marginal likelihood using the harmonic mean approach (Lartillot & Philippe, 2006; Xie, Lewis, Fan, Kuo, & Chen, 2011). We surmise that this finding may be attributed to the fact that when the priors are bounded (as we have done, but for other reasons) and does not permit extremely low probabilities to occur at the tails. Further, our emphasis here is to ask how the tool performs to select the true model and estimate parameters for prediction for a given analytical setup. And it is not about how accurately the tools estimate the actual value of the marginal likelihood. Thus, many of the criticisms (Lartillot & Philippe, 2006; Xie et al., 2011) may become irrelevant in practice, but this would require further enquiry.

As our simulation study shows, the two goals of model selection and parameter estimation cannot be simultaneously achieved under certain circumstances (scenarios 1–3), especially when information content is low (indicated by high RMSE values and correlation coefficients). Hence, researchers have to clearly prioritize their objectives prior to data analysis. If the goal, for example, is to find a model that best estimates population size N , then we recommend the use of Bayes factor (Gelfand–Dey with MAP approximation) or the Bayes factor (harmonic mean, due to its simplicity) because these appear to provide the most reliable estimates of N over all the simulation scenarios. However, if researchers are only interested to select the true

model, especially when data are less informative (scenarios 1–3), we recommend the posterior predictive loss approach since they favor the true model nearly a one-third of the times only in situations with such low information content. Of course, the dual objectives of model selection and parameter estimation are met when information content is moderate or high (scenarios 4–12), and as stated previously, we recommend either the Bayes factor (Gelfand-Dey with MAP approximation) or the Bayes factor (harmonic mean) in such cases. However, the posterior predictive loss (with the squared error loss function) as used here does not select the true model when information content is moderate to high.

We also do not recommend the use of the DICs or the WAICs, since they do not appear to outcompete other model selection tools, either from the standpoint of model selection or parameter estimation, in any of the simulation scenarios. This is an interesting finding, because tools such as the WAIC are much newer tools developed by applied Bayesians and recommended for hierarchical models (Gelman, Carlin, et al., 2014; Hooten & Hobbs, 2015). Thus, our study brings back focus on the need to assess the strength of inference from a model selection method by *conditioning* on a true model and consequently evaluating a *competing* set of model selection

methods prior to selecting the most appropriate one for the problem on hand. Table 7 summarizes the similarities and differences in inferences between our specific evaluation of model selection tools and the generally accepted evaluations. We believe this summary will motivate investigators to look deeper into their own models and the model selection methods they are using.

Our approach does not, strictly speaking, permit us to draw conclusions and make inferences on the most suitable model selection tools beyond the restrictive set of competing models and the settings we have used in this study. However, we demonstrate the applicability of the Bayes factor on a set of structurally very complex hierarchical models. This was possible by using our approach of partitioning the real-valued scalar and high-dimensional structural parameters. We anticipate that this general approach could address difficulties faced by ecologists attempting to use the Bayes factors for comparing very complex hierarchical models. We would expect that it would be much easier to use the Bayes factor for relatively simpler and a much larger class of hierarchical models (Kéry & Royle, 2015; Royle et al., 2013). We note with interest that it is also unclear whether the routinely used AIC works as an appropriate model selection tool for MLE-based SCR models as discussed in Efford and Mowat (2014). In real-world

TABLE 7 Contrasting the performance of model selection tools based on the intended purpose and perceived applicability with findings from our specific study. In the table, we provide answers to the following questions: (a) Does this approach select the true model? (b) Does this approach favor models providing reliable estimates of parameters (specifically for N)? (c) How difficult is the approach to implement in practice? Comments in bold draw attention to the noticeable differences between the expected performance of a tool and its performance in our particular study

Model selection tool	Intended purpose and applicability	Reference	Findings from our specific study
Bayes factor (by Gelfand-Dey estimator)	(a) Yes (b) Yes (c) Difficult		(a) Yes, very often (MAP) Yes, quite often (IL) (b) Yes, very often (MAP) Yes, quite often (IL) (c) Moderately difficult (MAP) Moderately difficult (IL)
Bayes factor (by harmonic mean estimator)	(a) Yes, but unreliable (b) Yes (c) Easy	Newton and Raftery (1994), and Kass and Raftery (1995)	(a) Yes, very often (b) Yes, very often (c) Easy
DIC1	(a) No (b) Yes (c) Easy	Spiegelhalter et al. (2002), Gelman, Hwang, & Vehtari, (2014), and Hooten and Hobbs (2015)	(a) No (b) No (c) Moderately difficult
DIC2	(a) No (b) Yes (c) Easy	Spiegelhalter et al. (2002), Gelman, Hwang, & Vehtari, (2014), and Hooten and Hobbs (2015)	(a) No (b) No (c) Moderately difficult (required MAP estimate)
WAIC1	(a) No (b) Yes (c) Easy	Watanabe (2010), Gelman, Hwang, & Vehtari, (2014), and Hooten and Hobbs (2015)	(a) No (b) No (c) Easy
WAIC2	(a) No (b) Yes (c) Easy	Watanabe (2010), Gelman, Hwang, & Vehtari, (2014), Hooten and Hobbs (2015)	(a) No, not often (b) Yes, but suboptimally (c) Easy
WAIC3	(a) No (b) Yes (c) Easy		(a) No (b) No (c) Easy
Posterior predictive loss	(a) No (b) Yes (c) Moderately difficult	Gelfand and Ghosh (1998), and Hooten and Hobbs (2015)	(a) No, not often (b) Yes, but suboptimally (c) Moderately difficult

problems, we can almost always find proper priors for parameters to ensure that Bayes factors exist. Therefore, we believe it is best to depend on tools built on sound statistical theory (by finding suitable approximations) rather than seeking answers based on seemingly elegant but not well-tested tools.

ACKNOWLEDGMENTS

We thank the Indian Statistical Institute for financial and administrative support. AMG thanks the Wildlife Conservation Society, New York, for partial funding support.

CONFLICT OF INTEREST

None declared.

AUTHORS CONTRIBUTION

SD, MD, and AMG conceived the ideas and designed methodology; SD conducted data analysis; SD, MD, and AMG drafted the article. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

R codes are available from GitHub, <https://github.com/soumenstat89/bmse.git>. Authors have used only simulated data sets in this article, which can be generated using the R codes provided. Hence, no particular data set is archived.

ORCID

Soumen Dey  <https://orcid.org/0000-0001-6270-2356>

REFERENCES

- Augustine, B. C., Royle, J. A., Kelly, M. J., Satter, C. B., Alonso, R. S., Boydston, E. E., & Crooks, K. R. (2018). Spatial capture-recapture with partial identity: An application to camera traps. *Annals of Applied Statistics*, *12*(1), 67–95.
- Berger, J. O., Ghosh, J. K., & Mukhopadhyay, N. (2003). Approximations and consistency of Bayes factors as model dimension grows. *Journal of Statistical Planning and Inference*, *112*(1), 241–258.
- Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. Chichester, UK: John Wiley & Sons.
- Bolker, B. M. (2008). *Ecological models and data in R*. Princeton, NJ: Princeton University Press.
- Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, *64*(2), 377–385.
- Broekhuis, F., & Gopalaswamy, A. M. (2016). Counting cats: Spatially explicit population estimates of cheetah (*acinonyx jubatus*) using unstructured sampling data. *PLoS ONE*, *11*(5), 1–15.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.
- Celeux, G., Forbes, F., Robert, C. P., & Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–673.
- Chan, J. C. C., & Eisenstat, E. (2015). Marginal likelihood estimation with the Cross-Entropy method. *Econometric Reviews*, *34*(3), 256–285.
- Clarke, J. L., Clarke, B., & Yu, C.-W. (2013). Prediction in M-complete problems with limited sample size. *Bayesian Analysis*, *8*(3), 647–690.
- Dass, S. C., & Lee, J. (2004). A note on the consistency of Bayes factors for testing point null versus non-parametric alternatives. *Journal of Statistical Planning and Inference*, *119*(1), 143–152.
- Dey, S., Delampady, M., Karanth, K. U., & Gopalaswamy, A. M. (2019). A spatially explicit capture-recapture model for partially identified individuals when trap detection rate is less than one. *Calcutta Statistical Association Bulletin*, 1–20. <https://doi.org/10.1177/0008068319837087>
- Efford, M. G., & Mowat, G. (2014). Compensatory heterogeneity in spatially explicit capture-recapture data. *Ecology*, *95*(5), 1341–1348.
- Elliot, N. B., & Gopalaswamy, A. M. (2017). Toward accurate and precise estimates of lion density. *Conservation Biology*, *31*(4), 934–943.
- Gardner, B., Sollmann, R., Kumar, N. S., Jathanna, D., & Karanth, K. U. (2018). State space and movement specification in open population spatial capture-recapture models. *Ecology and Evolution*, *8*(20), 10336–10344.
- Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*, 501–514. <https://doi.org/10.1111/j.2517-6161.1994.tb01996.x>
- Gelfand, A. E., & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, *85*(1), 1–11. <https://doi.org/10.1093/biomet/85.1.1>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. <https://doi.org/10.1007/s11222-013-9416-2>
- Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews*, *18*(1), 1–73.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis, theory and methods*. Springer texts in statistics. New York, NY: Springer.
- Ghosh, J. K., & Samanta, T. (2001). Model selection: An overview. *Current Science*, *80*(9), 1135–1144.
- Goldberg, J. F., Tempa, T., Norbu, N., Hebblewhite, M., Mills, L. S., Wangchuk, T. R., & Lukacs, P. (2015). Examining temporal sample scale and model choice with spatial capture-recapture models in the common leopard *Panthera pardus*. *PLoS ONE*, *10*(11), e0140757.
- Gopalaswamy, A. M., & Delampady, M. (2016). *Examining posterior propriety in the Bayesian analysis of capture-recapture models*. arXiv preprint arXiv:1611.02403.
- Hilborn, R., & Mangel, M. (1997). *The ecological detective: Confronting models with data (MPB-28)*. Princeton, NJ: Princeton University Press.
- Höge, M., Guthke, A., & Nowak, W. (2019). The hydrologists guide to Bayesian model selection, averaging and combination. *Journal of Hydrology*, *572*, 96–107.
- Höge, M., Wöhling, T., & Nowak, W. (2018). A primer for model selection: The decisive role of model complexity. *Water Resources Research*, *54*(3), 1688–1715.
- Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101–108.
- Karanth, K. U., & Nichols, J. D. (2017). *Methods for monitoring tiger and prey populations*. Singapore City, Singapore: Springer.

- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kéry, M., & Royle, J. A. (2015). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and static models*. Cambridge, MA: Academic Press.
- Lartillot, N., & Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology*, 55(2), 195–207.
- Lenk, P. (2009). Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *Journal of Computational and Graphical Statistics*, 18(4), 941–960.
- Millar, R. B. (2009). Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics*, 65(3), 962–969.
- Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1), 3–48.
- Robert, C. P. (2007). *The Bayesian choice, from decision-theoretic foundations to computational implementation* (2nd ed.). Springer texts in statistics. New York, NY: Springer.
- Royle, J. A. (2015). *Spatial capture-recapture with partial identity*. arXiv preprint arXiv:1503.06873.
- Royle, J. A., Chandler, R. B., Sollmann, R., & Gardner, B. (2013). *Spatial capture-recapture*. Cambridge, MA: Academic Press.
- Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Cambridge, MA: Academic Press.
- Royle, J. A., Karanth, K. U., Gopalaswamy, A. M., & Kumar, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, 90(11), 3233–3244.
- Royle, J. A., Sutherland, C., Fuller, A. K., & Sun, C. C. (2015). Likelihood analysis of spatial capture-recapture models for stratified or class structured populations. *Ecosphere*, 6(2), 1–11.
- Shibata, R. (1989). Statistical aspects of model selection. In J. C. Willems (Ed.), *From data to model* (pp. 215–240). Berlin, Heidelberg, Germany: Springer.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Sollmann, R., Furtado, M. M., Gardner, B., Hofer, H., Jácomo, A. T. A., Tôrres, N. M., & Silveira, L. (2011). Improving density estimates for elusive carnivores: Accounting for sex-specific detection and movements using spatial capture-recapture models for jaguars in central Brazil. *Biological Conservation*, 144(3), 1017–1024.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and t. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 64(4), 583–639.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 44–47. <https://doi.org/10.1111/j.2517-6161.1977.tb01603.x>
- Vehtari, A., & Ojanen, J. (2012). A survey of bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6, 142–228.
- Wang, L., & Meng, X. (2016). *Warp bridge sampling: The next generation*. arXivpreprint arXiv:1609.07690.
- Wang, Y., Chen, M., Kuo, L., & Lewis, P. O. (2018). A new Monte Carlo method for estimating marginal likelihoods. *Bayesian Analysis*, 13(2), 311.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec), 3571–3594.
- Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M. (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60(2), 150–160.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Dey S, Delampady M, Gopalaswamy AM. Bayesian model selection for spatial capture–recapture models. *Ecol Evol*. 2019;9:11569–11583. <https://doi.org/10.1002/ece3.5551>