Received: 7 September 2022 | Revised: 31 October 2022 | Accepted: 7 November 2022

DOI: 10.1002/ecy.3942

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Model selection in occupancy models: Inference versus prediction

Peter S. Stewart[1] | **Philip A. Stephens**[1] | **Russell A. Hill**[2] | **Mark J. Whittingham**[3] | **Wayne Dawson**[1]

[1]Department of Biosciences, Durham University, Durham, UK

[2]Department of Anthropology, Durham University, Durham, UK

[3]School of Natural and Environmental Sciences, Newcastle University, Newcastle-Upon-Tyne, UK

Correspondence
Peter S. Stewart
Email: peter.s.stewart@durham.ac.uk

Funding information
Natural Environment Research Council

Handling Editor: Brett T. McClintock

Abstract

Occupancy models are a vital tool for ecologists studying the patterns and drivers of species occurrence, but their use often involves selecting among models with different sets of occupancy and detection covariates. The information-theoretic approach, which employs information criteria such as Akaike's information criterion (AIC) is arguably the most popular approach for model selection in ecology and is often used for selecting occupancy models. However, the information-theoretic approach risks selecting models that produce inaccurate parameter estimates due to a phenomenon called collider bias, a type of confounding that can arise when adding explanatory variables to a model. Using simulations, we investigated the consequences of collider bias (using an illustrative example called M-bias) in the occupancy and detection processes of an occupancy model, and explored the implications for model selection using AIC and a common alternative, the Schwarz criterion (or Bayesian information criterion, BIC). We found that when M-bias was present in the occupancy process, AIC and BIC selected models that inaccurately estimated the effect of the focal occupancy covariate, while simultaneously producing more accurate predictions of the site-level occupancy probability than other models in the candidate set. In contrast, M-bias in the detection process did not impact the focal estimate; all models made accurate inferences, while the site-level predictions of the AIC/BIC-best model were slightly more accurate. Our results show that information criteria can be used to select occupancy covariates if the sole purpose of the model is prediction, but must be treated with more caution if the purpose is to understand how environmental variables affect occupancy. By contrast, detection covariates can usually be selected using information criteria regardless of the model's purpose. These findings illustrate the importance of distinguishing between the tasks of parameter inference and prediction in ecological modeling. Furthermore, our results underline concerns about the use of information criteria to compare different biological hypotheses in observational studies.

KEYWORDS
Akaike's information criterion, causal inference, collider bias, information-theoretic approach, model selection, occupancy models

INTRODUCTION

The patterns and drivers of species occurrence are of fundamental interest to ecologists. Predicting where species occur enables ecologists to tackle key problems such as understanding the spread of invasive species (Gormley et al., 2011), assessing the distributions of key species within protected areas (Midlane et al., 2014), and estimating the range size of populations and species to evaluate their extinction risk (Breiner & Bergamini, 2018) and recovery (Akçakaya et al., 2018). Understanding the drivers of occurrence is also important; interventions to mitigate the factors that threaten species must be informed by the diagnosis of those factors (Caughley, 1994). Many studies have aimed to infer how occurrence is driven by factors including forest degradation (Zimbres et al., 2018), wildfires (Hossack et al., 2013), and anthropogenic noise pollution (Chen & Koprowski, 2015).

A key challenge in studying species occurrence is that experimental manipulations of ecological systems may be physically impossible, logistically unfeasible, or unethical; consequently, ecologists are often constrained to the use of observational data. One approach to this challenge is to use a model that relates observed variation in species occurrence to one or more environmental covariates. The model can then be used to predict, or to explain (Shmueli, 2010): we can predict species occurrence at new sites, or examine the effect of each covariate to explain the drivers of occurrence. Occupancy models are often used because they deal with imperfect detection (MacKenzie et al., 2002). They do so by modeling the probability that a species occupying a site is detected, often including environmental covariates to explain variation in detectability among sites (MacKenzie et al., 2002). Occupancy models therefore contain one set of covariates for occupancy probability, and a second set for detection probability; the challenge is to select suitable sets of covariates to include in the model. This challenge can be framed as a problem of model selection (Burnham & Anderson, 2004; Forster, 2000; Johnson & Omland, 2004; Robins & Greenland, 1986).

The information-theoretic approach to model selection

The information-theoretic approach (Anderson et al., 2000; Burnham et al., 2011; Burnham & Anderson, 2001, 2004; Lukacs et al., 2007) compares models in terms of their relative Kullback–Leibler (KL) divergence, the relative distance between each model and "full reality," in units of information entropy (Burnham & Anderson, 2001; Forster, 2000; McElreath, 2021, p. 207). Information criteria, of which Akaike's information criterion (AIC; Akaike, 1973) is the most commonly used, estimate the relative KL divergence of each model using the sample data (McElreath, 2021, p. 219). AIC is calculated by taking the in-sample deviance (a measure of how well the model fits the data), and adding an overfitting penalty proportional to the number of parameters in the model (Akaike, 1973; Burnham et al., 2011). Consequently, AIC favors parsimonious models that balance underfitting and overfitting, with the aim of producing better out-of-sample predictions (McElreath, 2021, p. 192).

Causal inference

An alternative approach to model selection that has gained recent traction in ecology and evolution (e.g., Arif & MacNeil, 2022; Laubach et al., 2021) is causal inference. Causal inference is concerned with predicting the consequences of intervening in a system, as well as inferring counterfactual outcomes, events that might have happened, under hypothetical unrealized conditions (Pearl et al., 2016, p. 89). Importantly, causal inference is not about "inferring causation from correlation"; conclusions about causality cannot be made from the data alone, but require causal assumptions about the process that generated the data (Pearl et al., 2016, p. 5). To illustrate the key concepts and terminology of the causal inference approach we will discuss a hypothetical example, in which the goal is to infer how the density of an invasive plant affects the occupancy of a native animal (Figure 1).

In the causal inference approach, the first step is to employ subject expertise and the literature to identify variables that could be important in the system. This step closely resembles the "hard thinking" which is an essential part of the information-theoretic approach (Burnham et al., 2011). In our hypothetical example, we know of a native food plant that is regularly consumed by the animal species, and may therefore influence the animal's occupancy. Furthermore, there is evidence to suggest that the invasive plant tends to outcompete the native food

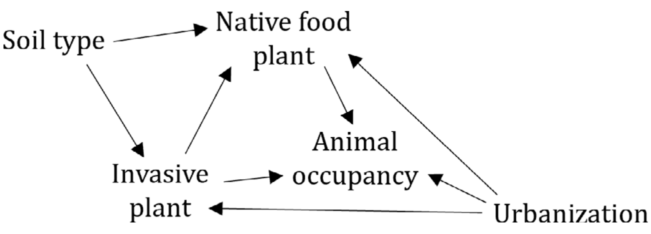

FIGURE 1 Directed acyclic graph for a hypothetical example in which we are interested in estimating the direct effect of the density of an invasive plant on animal occupancy.

plant, particularly on certain soil types. Finally, as our hypothetical study is conducted along an urban–rural gradient, the degree of urbanization is likely to be important.

The next step is to make assumptions about how these variables might be related to one another; these are known as "causal assumptions" (Pearl et al., 2016, p. 5). A key principle of causal inference is that these assumptions should be communicated clearly so that they are open to scrutiny, debate, sensitivity analysis, and verification (Pearl, 1995). Consequently, it is common to express causal assumptions graphically, usually as a directed acyclic graph (DAG; Pearl, 1995). In a DAG, variables are represented as nodes. The edges (arrows) linking each node represent the assumed mechanistic links between the variables (Greenland et al., 1999; Pearl, 1995). The sequence of edges linking one variable to another, regardless of which direction these edges are pointing in, is called a path (Pearl, 1995). In our example (Figure 1), we have assumed that the invasive plant, the native food plant, and urbanization all exert direct influences on the animal's occupancy. We have also assumed that the invasive plant affects the density of the native food plant through competition. Furthermore, we have assumed that urbanization also influences the densities of both the invasive and native plants. Finally, we have assumed that soil type does not influence occupancy directly, but that it does affect the densities of both the invasive and native plants.

Once we have specified a DAG, we must identify which effects we are interested in estimating. In our example, we could estimate the direct effect (invasive plant → animal occupancy) or the total effect (invasive plant → animal occupancy and invasive plant → native food plant → animal occupancy) of the invasive plant; in our example, the focal effect will be the direct effect. Once we have decided on a focal effect, we can analyze the DAG directly to identify a set of variables to condition on (i.e., include as covariates) that will allow us to estimate the effect. One strategy is to condition on the variables that satisfy the "back-door criterion," in which the aim is to "close all back-door paths" linking the focal explanatory and response variables (Pearl, 1995). A back-door path is defined as any path that has an arrow entering the focal explanatory variable (Pearl, 1995). Our example contains four back-door paths: (1) invasive plant ← soil type → native food plant → animal occupancy; (2) invasive plant ← soil type → native food plant ← urbanization → animal occupancy; (3) invasive plant ← urbanization → animal occupancy; (4) invasive plant ← urbanization → native food plant → animal occupancy.

Whether a path is open or closed depends on the direction in which arrows along the path are pointing. Paths that are a "fork" (e.g., $X \leftarrow Z \rightarrow Y$) or "pipe" (e.g., $X \rightarrow Z \rightarrow Y$) are open by default, and conditioning on the middle variable (Z) closes them (Greenland, 2003; McElreath, 2021, pp. 184–185; Pearl et al., 2016, p. 46). In contrast, paths that are a collider (e.g., $X \rightarrow Z \leftarrow Y$) are closed by default, and conditioning on the middle variable (Z) opens the path (Greenland, 2003; Greenland et al., 1999; McElreath, 2021, p. 185; Pearl et al., 2016, p. 46). A path with more than three variables only needs to be closed in one place to be closed overall (e.g., $X \leftarrow W \rightarrow Z \leftarrow Y$ is closed by the collider at Z).

In our example, the back-door paths 1 and 2 are closed by default because native food plant is a collider. However, as we are interested in the direct effect invasive plant → animal occupancy we need to close the indirect path, invasive plant → native food plant → animal occupancy, by conditioning on native food plant. This opens paths 1 and 2, but we can close both paths again by conditioning on either soil type or urbanization. If we condition on urbanization, then doing so also closes paths 3 and 4, meaning that all four back-door paths will be closed. Consequently, we can use the model: animal occupancy ~ invasive plant + native food plant + urbanization because it closes all of the back-door paths, satisfying the back-door criterion. We could also condition on soil type, but doing so is not required to estimate the direct effect. As DAG-based approaches are nonparametric in the sense that the forms of the functions represented by edges do not have to be specified (Greenland et al., 1999; Pearl, 1995), we would also be free to incorporate linear interactions between these covariates, or model their effects as nonlinear functions.

Finally, we can explore the consequences of changing the assumptions embodied in our DAG, to see whether our inferences hold under different sets of assumptions. For instance, we could ask "what if urbanization does not affect the density of the invasive plant?", remove the arrow urbanization → invasive plant, and re-analyze the DAG. Doing so, we see that our model still satisfies the back-door criterion; our conclusions are robust to altering this assumption. We can also modify the DAG to answer questions such as "what if there was an unmeasured confounding variable affecting the densities of both the invasive and native plants?". By adding a new variable invasive plant ← unmeasured variable → native food plant and re-analyzing the DAG, we can see again that the same model structure is supported because it satisfies the back-door criterion, and thus our conclusions still hold. Where a modified DAG supports a different model structure, we can run the new model and compare the effect estimates with those of the original model.

Collider bias and the information-theoretic approach

Proponents of the information-theoretic approach have argued that each model in the candidate set should represent a different biological hypothesis, and that the models' relative AIC scores indicate the strength of evidence for each hypothesis (Burnham et al., 2011). However, insights from causal inference reveal a potential problem: collider bias. Collider bias arises when back-door paths are opened due to conditioning on collider variables (Greenland, 2003), and is a form of included-variable bias or "bad control" (Cinelli et al., 2022). This is in contrast with the classical notion of confounding (Figure 2A), which is a form of omitted variable bias (Clarke, 2005). As collider covariates and classical confounds exhibit a similar degree of correlation to the focal explanatory variable (Figure 2D), and these correlations may be masked or otherwise distorted by the action of other variables or nonlinear relationships between the covariates, it is not possible to avoid collider bias by checking the explanatory variables for multicollinearity.

As AIC and other information criteria select models based on their expected predictive performance, they are vulnerable to collider bias: including a collider covariate tends to improve a model's AIC score, while simultaneously resulting in an estimated effect that is far from the true value (Figure 2A–C; Luque-Fernandez et al., 2019). Consequently, recent studies have argued that it is essential to consider whether the purpose of a model is inference (i.e., explanation) or prediction when deciding on a model selection strategy (e.g., Arif & MacNeil, 2022; Laubach et al., 2021). However, the implications for models like occupancy models, which contain multiple submodels, are unclear.

To address this topic, we investigated the consequences of a form of collider bias (using an illustrative example known as "M-bias"; Cinelli et al., 2022; Greenland, 2003) in an occupancy modeling framework, and explored the implications for model selection using the information-theoretic approach (using AIC). We also examined the performance of a common alternative to AIC, the Schwarz criterion (or Bayesian information criterion, BIC; Schwarz, 1978). BIC is built upon different philosophical foundations to AIC, and is not based upon information theory (Johnson & Omland, 2004); some authors have suggested BIC can be used for selecting the "true" model from the candidate set (Aho et al., 2014). In our simulation-based approach, we generated datasets where M-bias was present in the occupancy process, the detection process, or both. We then fitted occupancy models with different sets of covariates to these datasets, and evaluated them on the accuracy of their parameter inferences, the accuracy of their site-level occupancy predictions, and their level of support from AIC and BIC.

METHODS

M-bias as an illustrative example

M-bias is a common illustrative example in the causal inference literature (e.g., Cinelli et al., 2022; Greenland, 2003), in which an "M"-shaped back-door path (e.g., Figure 3A, left panel) is opened by conditioning on the collider variable (D in Figure 3A), confounding the estimate of the focal effect ($X \rightarrow \psi$ in Figure 3A). When the back-door path contains latent (unobserved) variables (A and C in Figure 3A), it is impossible to condition on them to close the path because they are unobserved, meaning the correct approach is to not condition on the collider.

Simulation study

To explore the effects of M-bias in both the occupancy and detection components of an occupancy model, we simulated three different scenarios (Figure 3) in which the focal effect was the effect of variable X on occupancy probability (ψ). In the first scenario (Figure 3A), ψ was part of an M-graph while the detection probability (p) was fixed at 0.5. In the second scenario (Figure 3B), ψ depended only on X, and p was now part of an M-graph. In the final scenario (Figure 3C), both the occupancy and detection probabilities were part of M-graphs.

All three simulations followed the same process: (1) generate a dataset with known parameter values, using the relationships between variables embodied in the relevant DAGs (Figure 3); (2) fit a number of occupancy models to the dataset (Figure 3); (3) evaluate each model's accuracy in parameter estimation and prediction; (4) evaluate each model's quality under the

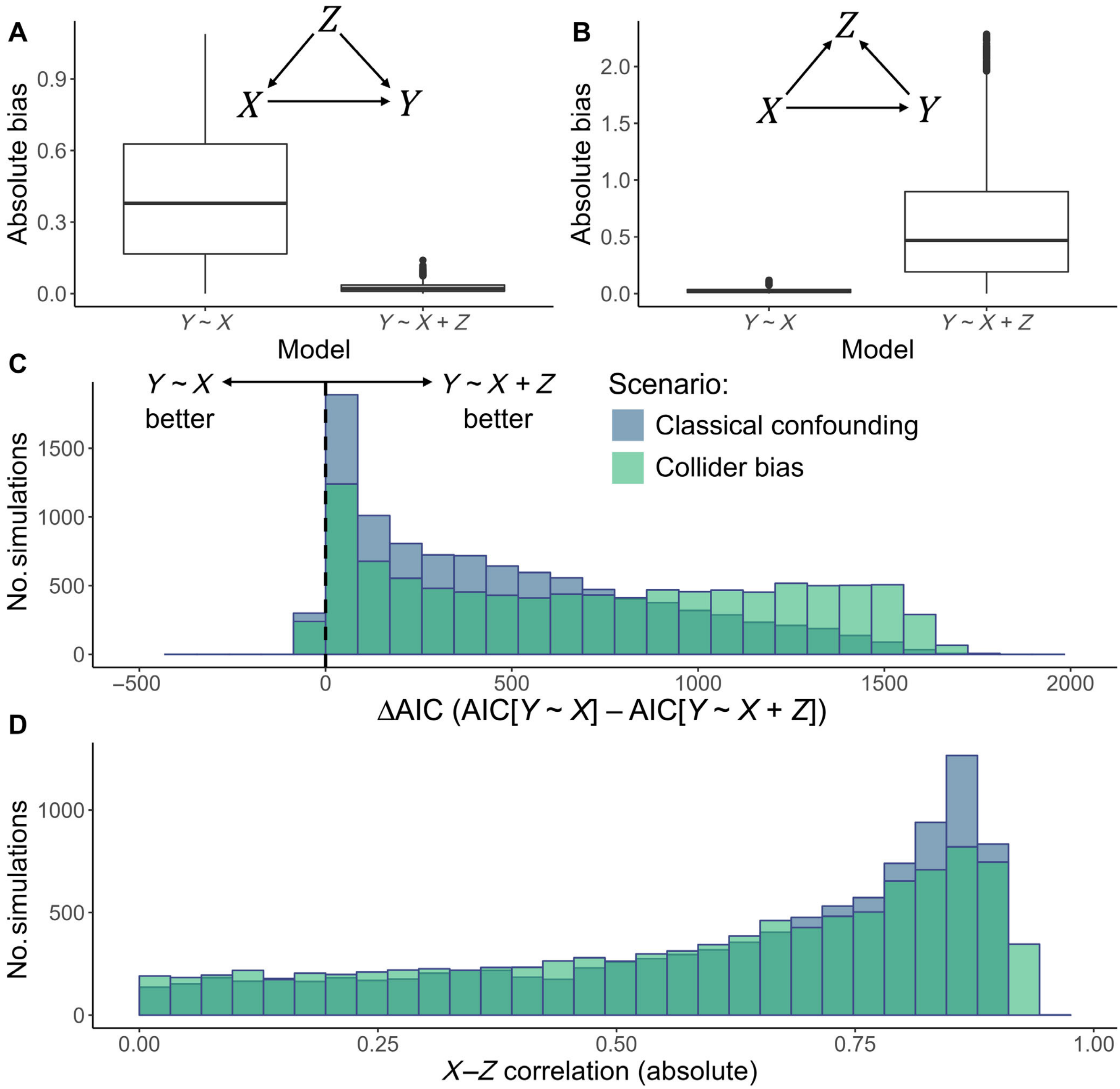

FIGURE 2 Luque-Fernandez et al. (2019) presented simulations illustrating classical confounding and collider bias in a linear model. We extended their example by conducting 10,000 iterations for each example, using effect sizes drawn from a uniform distribution between −2 and 2. In classical confounding (A), including the variable *Z* reduces the absolute bias when estimating the effect of *X* on *Y*. Conversely, in the collider example (B), including *Z* increases the absolute bias. However, in both cases Akaike's information criterion (AIC) favors the model that includes *Z* (C), illustrating that AIC does not always favor models that produce accurate parameter estimates. Furthermore, the absolute correlation between *X* and *Z* is similar in both scenarios (D), meaning that checking for multicollinearity cannot reliably help to select the model that estimates β_{XY} more accurately, and that adding highly collinear explanatory variables can sometimes improve inferential accuracy. Code to reproduce the simulations is available in Stewart (2022) at https://doi.org/10.5281/zenodo.7043335.

information-theoretic framework. Each simulation was repeated 1000 times. We conducted our simulations in R (v 4.0.5; R Core Team, 2021), and provided code to reproduce our simulations and analyses in Stewart (2022) at https://doi.org/10.5281/zenodo.7043335.

Generating a dataset

Data were simulated for 3000 sites with 40 surveys each. The number of sites was deliberately high to ensure that any inaccuracy was not primarily driven by an

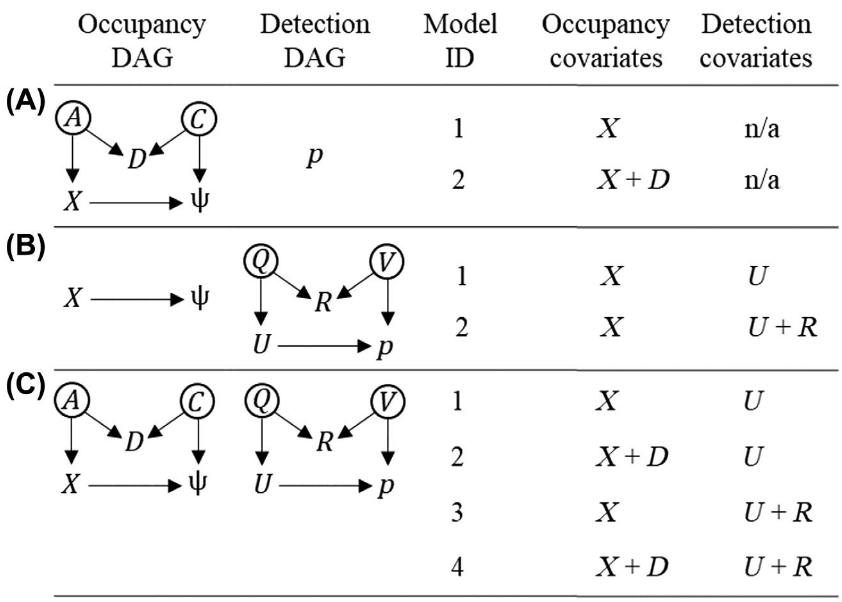

FIGURE 3 Data-generating processes and model structures for the three scenarios: (A) M-bias in the occupancy process, (B) M-bias in the detection process, (C) M-bias in both the occupancy and detection processes. Data-generating processes are represented as directed acyclic graphs (DAGs). ψ is the occupancy probability, p is the detection probability, X *is* the focal explanatory variable, and A, C, D, Q, U, R, and V represent other explanatory variables. Circled variables are latent. All models included intercept terms for occupancy and detection.

underpowered design. Repeating the simulations with two simulated surveys yielded qualitatively similar results (Appendix S1: Figures S7–S9).

To generate the data, we first drew effect sizes for each arrow in the DAG from a uniform distribution ($\min = -1$, $\max = 1$). Values for explanatory variables with no ingoing arrows on the DAG were then drawn from a normal distribution ($\text{mean} = 0$, $\text{SD} = 1$). We then generated values for the other explanatory variables from the appropriate variables and effect sizes (i.e., those from ingoing arrows on the DAG; Figure 3), plus a "disturbance term" (sensu Pearl, 1995) drawn from a normal distribution ($\text{mean} = 0$, $\text{SD} = 0.025$). We then calculated the log-odds of occupancy and detection as a linear combination of the effect sizes and explanatory variables with ingoing arrows on the DAG (Figure 3), and took the inverse-logit to obtain the probability. The true occupancy state of each site was then simulated as a Bernoulli trial with the probability of success equal to the occupancy probability. Finally, detection histories for each site were generated as a sequence of Bernoulli trials, with the probability of success equal to the true occupancy state multiplied by the detection probability.

Fitting models

Occupancy models were fitted to each dataset using the *occu* function in the R package unmarked (v.1.0.0; Fiske & Chandler, 2011), which implements the single-season occupancy model developed by MacKenzie et al. (2002). The models used the logit link function. We fitted models with various combinations of observed variables (i.e., excluding latent variables) for each scenario (Figure 3).

Evaluating model performance

In each scenario, all models were evaluated on the accuracy of their parameter inferences and predictions. To quantify how accurately each model estimated the effect of covariate X on the occupancy probability ψ, we calculated the bias and absolute bias:

$$\text{Bias} = \widehat{\beta}_{X\psi} - \beta_{X\psi},$$

$$\text{Absolute Bias} = |\widehat{\beta}_{X\psi} - \beta_{X\psi}|,$$

where $\widehat{\beta}_{X\psi}$ and $\beta_{X\psi}$ are the estimated and true effects of X on ψ, respectively. Additionally, we checked whether the true value, $\beta_{X\psi}$, was found within the 95% CI surrounding the $\widehat{\beta}_{X\psi}$ estimate, and then checked whether the sign (positive or negative) of $\widehat{\beta}_{X\psi}$ was the same as that of $\beta_{X\psi}$.

To evaluate each model's predictions, we used the predict function in R to predict the occupancy probability value for each site in two datasets. We first made predictions for the data to which the model was fitted, to examine how the model retrodicted the sample. We then examined the model's performance in out-of-sample prediction by making predictions for a new dataset (also 3000 sites), which was generated using the same true parameter values as the original dataset. To assess the accuracy of the model retrodictions and predictions, we calculated the mean absolute error:

$$\text{Mean absolute error} = \frac{1}{n}\sum_{i=1}^{n}|\widehat{\psi}_{\rm i} - \psi|,$$

where $\widehat{\psi}_i$ and ψ_i are the estimated and true occupancy probabilities for site i, respectively, and n is the number of sites. Additionally, we calculated the proportion of sites for which the true occupancy probability was within the prediction's 95% CI.

Evaluating models under the information-theoretic framework

To examine the degree of support for each model under the information-theoretic framework we obtained the AIC value for each model from the model's summary table. Proponents of the information-theoretic approach have advocated for multimodel inference (e.g., Burnham & Anderson, 2004), in which inferences are made using the entire candidate set of models, each weighted using Akaike weights derived from AIC. We calculated Akaike weights (w) for each model m as:

$$w_m = \frac{\exp(-0.5 \times \Delta\text{AIC}_m)}{\sum_{r=1}^{R}\exp(-0.5 \times \Delta\text{AIC}_r)},$$

where ΔAIC_m is the difference between the AIC of model m and the lowest AIC value for the set of models in the scenario, and R is the number of models in the scenario.

We also considered BIC (Schwarz, 1978) as an alternative to AIC. We calculated BIC and BIC weights for each model using the R package *AICcmodavg* (Mazerolle, 2020).

RESULTS

Scenario 1: M-bias in the occupancy process

When M-bias was present in the occupancy process, model 1 ($\psi \sim X$) estimated the true effect of X on ψ much more accurately than model 2 ($\psi \sim X + D$) (Figure 4A,B; Appendix S1: Table S1). However, comparing the models' predictive accuracy showed the opposite picture; model 1 generally produced worse predictions than model 2 (Figure 5A,B; Appendix S1: Table S1), and similar results were observed for retrodictive accuracy (Appendix S1: Figure S1). AIC and BIC both showed clear support for model 2 in the majority of simulations (Figure 6A,B); in 80.4% of simulations model 2 received an Akaike weight of >0.99, and in 52.2% of simulations it received the entire weight (Figure 6B). The few simulations in which model 1 received more weight were mostly those in which $\beta_{C\psi}$ was small (Appendix S1: Figure S2). A similar pattern of results was observed for BIC (Appendix S1: Figure S2), although when BIC assigned weight to model 1 it generally assigned more weight than AIC (Figure 6A).

Scenario 2: M-bias in the detection process

When M-bias was present in the detection process, both models 1 ($\psi \sim X,\ p \sim U$) and 2 ($\psi \sim X,\ p \sim U + R$) accurately estimated the effect of X on ψ (Figure 4C,D; Appendix S1: Table S1). Both models also made accurate predictions, although those of model 2 were more accurate (Figure 5C,D; Appendix S1: Table S1). Similar results were observed for retrodictive accuracy (Appendix S1: Figure S1C,D). Both AIC and BIC assigned more weight to model 2 in most simulations (Figure 6C,D).

Scenario 3: M-bias in the occupancy and detection processes

When M-bias was present in both the occupancy and detection processes, models 1 ($\psi \sim X,\ p \sim U$) and 3 ($\psi \sim X,\ P \sim U + R$) estimated the effect of X on ψ much more accurately than models 2 ($\psi \sim X + D,\ P \sim U$) and 4 ($\psi \sim X + D,\ P \sim U + R$) (Figure 4E–H; Appendix S1: Table S1). In general, the 95% confidence interval around the estimate in models 2 and 4 only contained the true value when β_{AD} and $\beta_{C\psi}$ (and to a lesser extent β_{AX}) were relatively small (Appendix S1: Figures S4, S5). In contrast, models 2 and 4 made more accurate predictions than models 1 and 3 (Figure 5E–H; Appendix S1: Table S1), and similar results were obtained for retrodictive accuracy (Appendix S1: Figure S1E–H). Both AIC and BIC showed clear support for model 4 in the majority of simulations (Figure 6H); the model received an Akaike weight of >0.99 in 63.0% of the simulations. While model 3 did occasionally receive weight, this mostly occurred when $\beta_{C\psi}$ was small (Appendix S1: Figure S6) and it still never received the entire weight (Figure 6G).

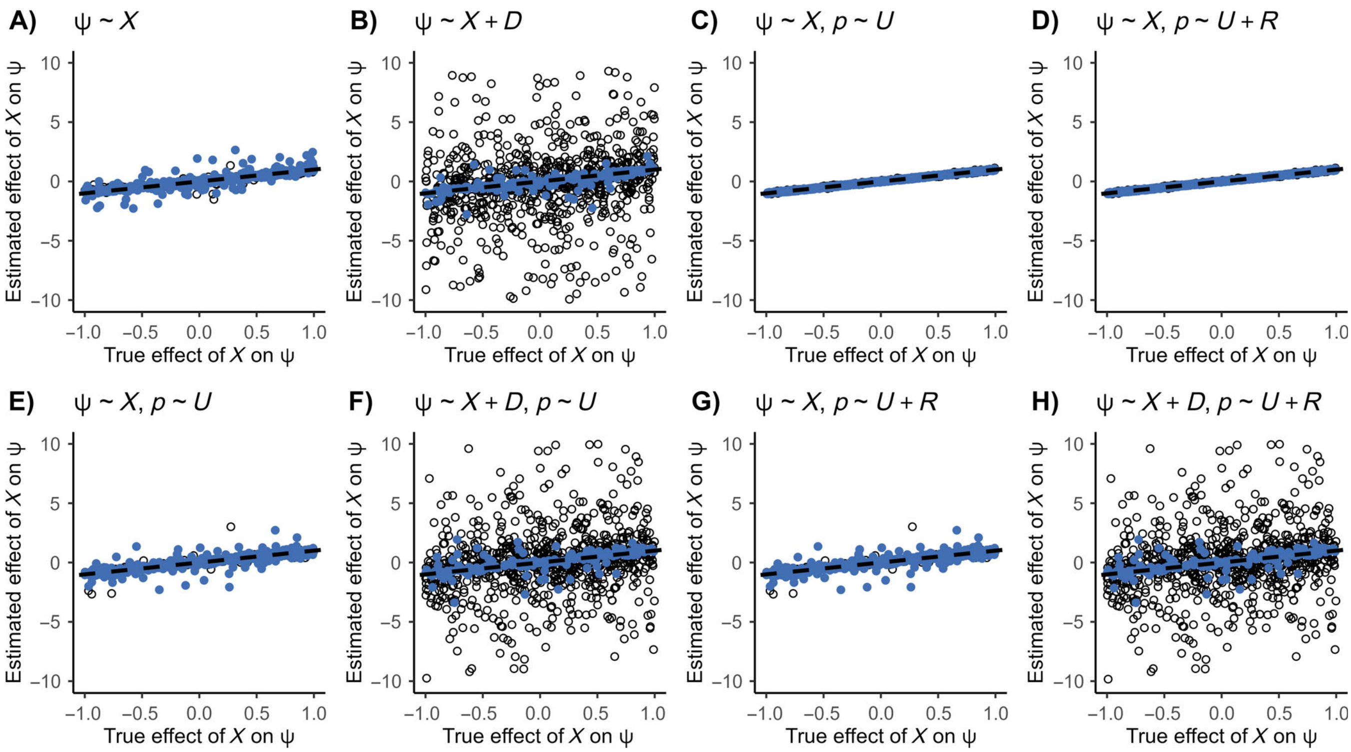

FIGURE 4 True versus estimated effect of X on occupancy probability (ψ), for the following occupancy models: (A) scenario 1, model 1; (B) scenario 1, model 2; (C) scenario 2, model 1; (D) scenario 2, model 2; (E) scenario 3, model 1; (F) scenario 3, model 2; (G) scenario 3, model 3; (H) scenario 4, model 4. Each point represents the result from one simulation, with 1000 simulations in total. The y-axis is truncated at -10 and 10; plots (B), (F), (G) and (H) omit 34, 33, 1, and 45 points respectively that lay outside this range. Blue points indicate that the true value was contained within the estimate's 95% CI, while unfilled circles indicate that the true value was not contained within the interval. The dashed black line indicates equality between the true and estimated effect. Each model's covariates for ψ and the detection probability (p) are shown above their respective plot.

BIC weights were similar to the Akaike weights, although BIC assigned more weight to model 3 in some simulations (Figure 4G), again generally when $\beta_{C\psi}$ was small (Appendix S1: Figure S6).

DISCUSSION

We investigated the consequences of M-bias (a specific form of collider bias) for occupancy modeling, and explored the implications for model selection using AIC and BIC. In our simulations, we observed that when M-bias was present in the occupancy process, AIC and BIC favored a model that produced a highly inaccurate estimate of the focal effect but produced more accurate predictions and retrodictions of the site-level occupancy probability. This reflects the fact that AIC and BIC aim to select models that produce better out-of-sample predictions (McElreath, 2021, p. 192). In contrast, M-bias in the detection process did not result in inaccurate estimates of the focal effect. However, the AIC/BIC-best models made better predictions and retrodictions. We observed the same results when M-bias was present in both the occupancy and detection processes: the model favored by AIC and BIC produced inaccurate inferences but more accurate predictions, while models made similarly accurate inferences regardless of M-bias in the detection process. These results have important implications for model selection in occupancy models, as well as for how the information-theoretic approach is applied in ecological modeling more generally.

Information criteria select models that produce poor parameter inferences, but good predictions

When M-bias was present in the occupancy process, the model that received the greatest support from AIC and BIC produced highly inaccurate estimates of the effect of the variable of interest (X) on the occupancy probability (ψ). The models that received the majority of the AIC and BIC weight were only able to estimate the direction of the focal effect correctly in 65.8% of cases at best, little better than

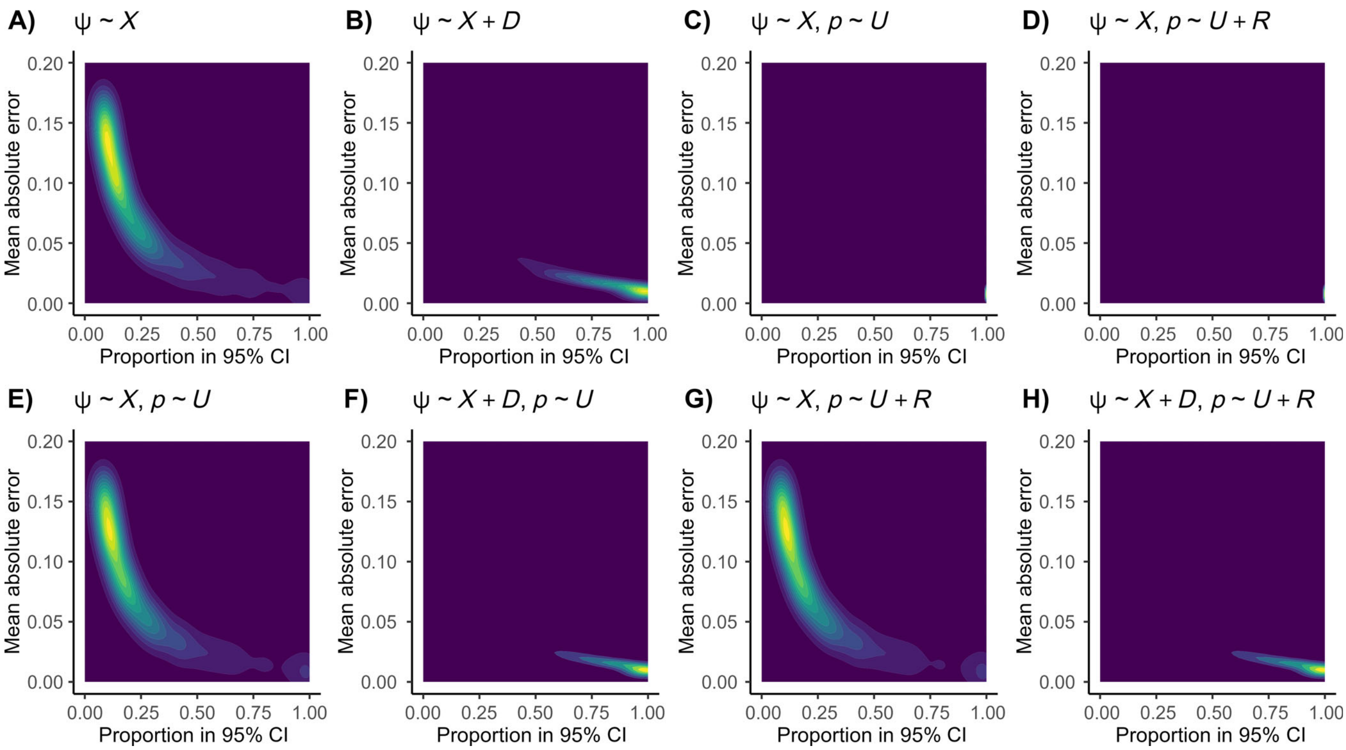

FIGURE 5 Kernel density estimate contours showing two measures of predictive accuracy when predicting site-level occupancy probability (ψ), for 1000 simulations. The *x*-axis shows the proportion of sites (out of 3000) for which the 95% CI around the model's prediction contained the true occupancy probability. The *y*-axis shows the mean absolute error. Thus, the bottom right of each plot indicates higher predictive accuracy, while the top left indicates lower predictive accuracy. The density of simulations is shown by the contours, with lighter colors indicating a higher density of simulations. Results are displayed for the following occupancy models: (A) scenario 1, model 1; (B) scenario 1, model 2; (C) scenario 2, model 1; (D) scenario 2, model 2; (E) scenario 3, model 1; (F) scenario 3, model 2; (G) scenario 3, model 3; (H) scenario 4, model 4. Each model's covariates for ψ and the detection probability (*p*) are shown above their respective plot.

the accuracy we would expect from guessing. Such biased estimates are not informative about the drivers underlying the observed pattern, nor do they accurately predict the consequences of intervening in the system; in a real conservation problem in which decisions are informed by occupancy models (e.g., Chen & Koprowski, 2015; Hossack et al., 2013; Zimbres et al., 2018), the results could be disastrous.

While the models supported by AIC and BIC produced biased parameter estimates, they also produced more accurate predictions and retrodictions of the occupancy probability at each site. This is because these models include the variable *D* that has an open path to ψ; including *D* provides additional information about the variation in ψ, improving prediction. From the perspective of AIC and BIC, including *D* results in a reduced in-sample deviance that typically outweighs the penalty for adding the additional variable; this reduction must be greater to outweigh BIC's larger penalty term, which is why BIC was more conservative in its tendency to select confounded models in our simulations (Figure 6). This also explains why AIC and BIC gave more weight to the nonconfounded model (omitting *D*) when $\beta_{C\psi}$ was close to zero (Appendix S1: Figures S2, S6); the near-zero effect of *C* on ψ meant that the path from *D* to ψ through *C* was almost blocked (the other path from *D* to ψ was blocked by conditioning on *X*), and therefore *D* explained relatively little variation in ψ.

In contrast with the effects of M-bias in the occupancy process, M-bias in the detection process did not affect inferences about the effect of *X* on ψ. Additionally, including the collider variable *R* in the detection submodel improved the accuracy of the model's predictions of the site-level occupancy probability. These results can again be explained by considering how the path structure between variables will affect the change in deviance when a variable is included; as the variable *R* has an open path to *p*, including *R* explains additional variation in the detection probability, reducing the deviance and allowing the model to better account for imperfect detection when estimating the occupancy probability. As the detection probability is generally regarded as a nuisance parameter (Karavarsamis, 2015), it is inconsequential that the effect of the other detection covariate (*U*) will be

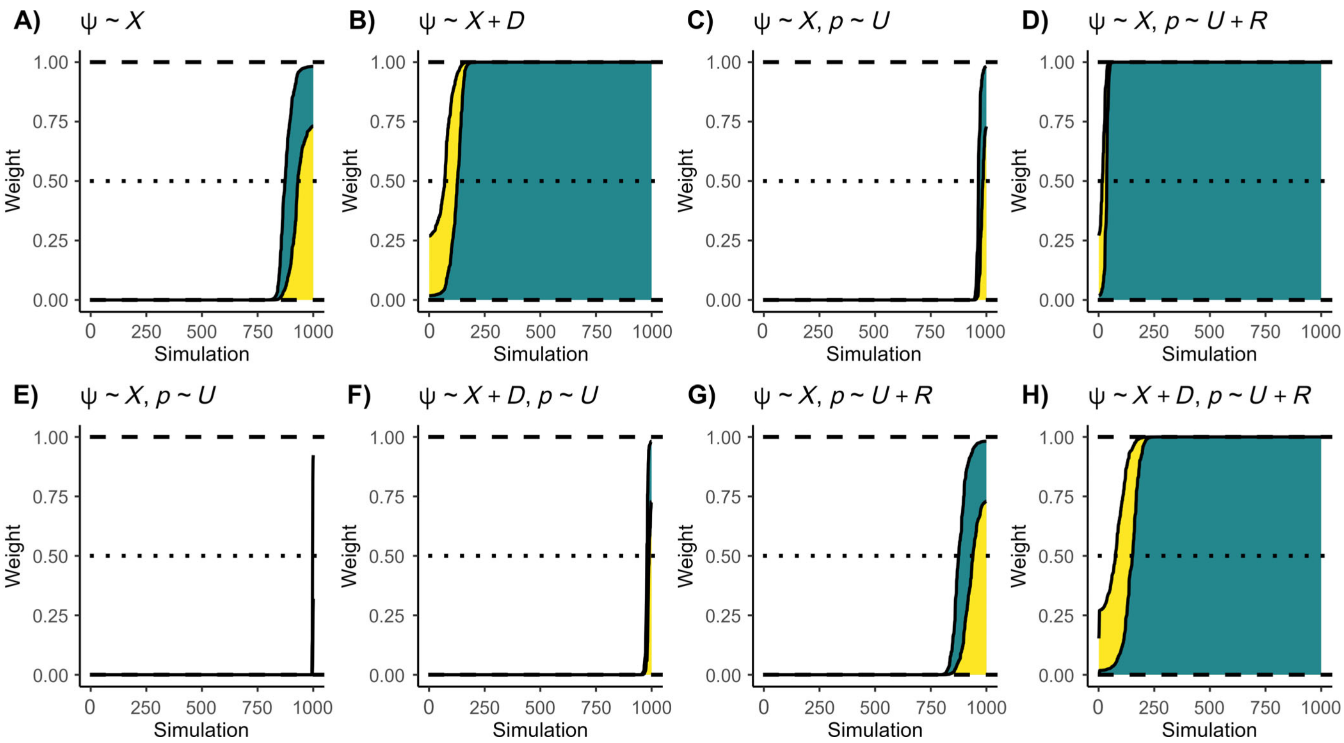

FIGURE 6 Akaike weight (yellow area) and Bayesian information criterion (BIC) weight (blue area) for 1000 simulations of eight occupancy models. Simulations are shown ranked by weight, with higher Akaike and BIC weights shown on the right. The panels display: (A) scenario 1, model 1; (B) scenario 1, model 2; (C) scenario 2, model 1; (D) scenario 2, model 2; (E) scenario 3, model 1; (F) scenario 3, model 2; (G) scenario 3, model 3; (H) scenario 4, model 4. Each model's covariates for the occupancy probability (ψ) and the detection probability (p) are shown above their respective plot. Dashed horizontal lines are shown for weights of 0, 0.5, and 1.

confounded. Therefore, information criteria can be used for selecting detection covariates.

The tendency for information criteria to favor confounded models with greater predictive ability is not confined to collider bias. For example, simulations by McElreath (2021, pp. 226–228) showed that information criteria tend to select models that condition on the mediator (M) in a pipe (e.g., $X \rightarrow M \rightarrow \psi$), inducing post-treatment bias (Rosenbaum, 1984). This occurs because adding M explains additional variation in ψ, while also blocking the causal path that runs from X to ψ (McElreath, 2021, p. 228). We also expect these results to apply in other scenarios, such as case–control bias (Cinelli et al., 2022). Finally, the M-bias example illustrates that latent variables can result in information criteria favoring confounded models, and hence that considering these variables is critical when drawing inferences.

Inference and prediction are separate tasks

The key point supported by our results is that inference and prediction are separate tasks that should not be conflated in model selection (Laubach et al., 2021; McElreath, 2021, p. 226; Shmueli, 2010). We echo Gelman and Rubin's (1995) criticism of selecting "a model that is adequate for specific purposes without consideration of those purposes." In the context of occupancy models, both explanation and prediction are important objectives, and conflating the two does justice to neither. Furthermore, our results emphasize the importance of considering not only the model's purpose, but also the purpose of submodels within the model; the purpose of the occupancy submodel depends on whether we are interested in predicting the occupancy state or inferring its drivers, while the detection submodel's purpose is usually prediction of the detection probability. Consequently, how occupancy covariates are chosen depends on the purpose of the model—information criteria are suitable if the purpose of the model is prediction, but are unlikely to be if the purpose is parameter inference—while detection covariates can generally be selected using information criteria. This advice also applies to cross-validation; the choice of model made by AIC is asymptotically equivalent to that made by leave-one-out cross-validation (Stone, 1977).

Using information criteria to compare biological hypotheses in observational studies is risky

The importance of distinguishing between inference and prediction has wider implications for how information-theoretic model selection is applied in ecology. Proponents of the information-theoretic approach have argued that it is possible to compare multiple a priori specified models, each representing a different biological hypothesis, with the relative AIC scores indicating the strength of evidence for each hypothesis (Burnham et al., 2011; Johnson & Omland, 2004; Richards, 2005). However, using information criteria in this way conflates inference and prediction; information criteria select models that make better predictions, but these same models can contain spurious effect sizes that hold no biological meaning, while the effects of biologically important covariates are confounded. This is not only the case for occupancy models; the occupancy models we used are just an extension of logistic regression (Clark & Altwegg, 2019), and these points apply to other forms of linear model as well (Luque-Fernandez et al., 2019; McElreath, 2021, pp. 226–228). The implication is that using information-theoretic model selection to compare biological hypotheses in observational studies carries substantial risks.

The information-theoretic approach and causal inference are complementary

While we argue that comparing biological hypotheses using the information-theoretic approach is risky, and that we prefer a causal inference-based approach for this purpose, we must emphasize that we are not arguing that the information-theoretic approach is flawed or useless for model selection. Information criteria select models from the "predictive point of view" (Akaike, 1998), while causal inference is concerned with estimating the effects of covariates, so we see the two approaches as complementary. In the case of occupancy models the two approaches may be used side by side in a single analysis, in which occupancy covariates are chosen based on causal assumptions embodied in a DAG, while the detection covariates are selected using the information-theoretic approach.

We also argue that causal inference and the information-theoretic approach are complementary because they share philosophical underpinnings. In the information-theoretic approach, it is vital to employ subject expertise and "hard thinking" to develop hypotheses that are compared as models (Burnham et al., 2011; Lukacs et al., 2007); in causal inference, subject expertise and a priori thought are vital in making the causal assumptions that are embodied in the DAG (Greenland et al., 1999; Pearl, 1995). Causal inference thus provides a framework to support the "hard thinking" required in ecological modeling (Grace & Irvine, 2020). Proponents of the information-theoretic approach also recognize that "a proper analysis must consider the science context and cannot successfully be based on 'just the numbers'" (Burnham & Anderson, 2004). Similarly, proponents of causal inference argue that conclusions cannot be drawn from the data alone, but require causal assumptions that come from the scientific context of the model (Pearl et al., 2016, p. 5).

Another feature of the information-theoretic approach is that Chamberlin's (1890) method of multiple working hypotheses is often emphasized (Burnham & Anderson, 2004; Elliott & Brook, 2007). We argue that causal inference is very compatible with Chamberlin's method; constructing a causal model forces us to consider multiple explanations for a phenomenon, guarding against the threat of "parental affection for a favorite theory" that concerned Chamberlin. Due to the relatively static nature of causal models, we argue they are especially suited to the case of multiple working hypotheses in parallel (Elliott & Brook, 2007), in which causation operates through multiple factors simultaneously. Moreover, the tools of causal inference allow this parallel case to be extended to more complex situations with indirect effects, rather than constraining our thinking to simple additive terms and interactions.

A caveat: Model selection is more than selecting covariates

We have focused on the choice of covariates, which is a key aspect of model selection, but another vital part of model selection is selecting specific mathematical functions to relate these variables to one another (Johnson & Omland, 2004). However, as the rules of causal inference are nonparametric (Greenland et al., 1999; Pearl, 1995) our conclusions hold irrespective of what functional forms are chosen, and we consider any role of information criteria in selecting these functions to be beyond the scope of our article.

Summary

We have demonstrated that when a form of collider bias known as M-bias is present in the occupancy process, occupancy models that are favored by AIC and BIC produce inaccurate parameter estimates but accurate predictions. In contrast, M-bias in the detection process does not affect the accuracy of parameter estimates. The key conclusion supported by these results is that inference and prediction are separate tasks that should not be conflated during model selection. The correct choice of model selection procedure depends on the purpose for which the occupancy model will be used. Information-theoretic approaches are suitable for selecting occupancy covariates if the model is to be used for predicting the site-level occupancy probability. However, if the goal is instead to infer the effect of environmental covariates on occupancy, then the use of information criteria carries significant risks; we advocate for an approach based on causal inference in this situation. Our results support the use of information-theoretic methods to select detection covariates regardless of the model's purpose, as long as detection probability is treated as a nuisance parameter. As single-season occupancy models are in essence a form of logistic regression, our results have wider implications for the use of information-theoretic model selection in ecology. In particular, we argue that our results, alongside those of others (Arif & MacNeil, 2022; Luque-Fernandez et al., 2019; McElreath, 2021), underscore the risks associated with using the information-theoretic approach to compare biological hypotheses in observational studies. Causal inference and the information-theoretic approach share similar philosophical underpinnings, and should be seen as complementary tools that accomplish different tasks.

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers and the editor for their helpful and insightful comments on our manuscript, and the members of the Conservation Ecology Group (Durham University) for their stimulating discussions. Peter S. Stewart was supported by the Natural Environment Research Council (NERC) through the Iapetus2 Doctoral Training Partnership.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

All data and code (Stewart, 2022) are available in Zenodo at https://doi.org/10.5281/zenodo.7043335.

ORCID

Peter S. Stewart https://orcid.org/0000-0001-6338-6279
Philip A. Stephens https://orcid.org/0000-0001-5849-788X
Russell A. Hill https://orcid.org/0000-0002-7601-5802
Mark J. Whittingham https://orcid.org/0000-0003-4362-3454
Wayne Dawson https://orcid.org/0000-0003-3402-0774

REFERENCES

Aho, K., D. Derryberry, and T. Peterson. 2014. "Model Selection for Ecologists: The Worldviews of AIC and BIC." *Ecology* 95: 631–6.

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Proceeding of the Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Kaski, 267–81. Budapest: Akademiai Kiado.

Akaike, H. 1998. "Prediction and Entropy." In *Selected Papers of Hirotugu Akaike, Springer Series in Statistics*, edited by E. Parzen, K. Tanabe, and G. Kitagawa, 387–410. New York, NY: Springer.

Akçakaya, H. R., E. L. Bennett, T. M. Brooks, M. K. Grace, A. Heath, S. Hedges, C. Hilton-Taylor, et al. 2018. "Quantifying Species Recovery and Conservation Success to Develop an IUCN Green List of Species." *Conservation Biology* 32: 1128–38.

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. "Null Hypothesis Testing: Problems, Prevalence, and an Alternative." *The Journal of Wildlife Management* 64: 912–23.

Arif, S., and M. A. MacNeil. 2022. "Predictive Models aren't for Causal Inference." *Ecology Letters* 25: 1741–5.

Breiner, F. T., and A. Bergamini. 2018. "Improving the Estimation of Area of Occupancy for IUCN Red List Assessments by Using a Circular Buffer Approach." *Biodiversity and Conservation* 27: 2443–8.

Burnham, K. P., and D. R. Anderson. 2001. "Kullback–Leibler Information as a Basis for Strong Inference in Ecological Studies." *Wildlife Research.* 28: 111.

Burnham, K. P., and D. R. Anderson. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods & Research* 33: 261–304.

Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons." *Behavioral Ecology and Sociobiology* 65: 23–35.

Caughley, G. 1994. "Directions in Conservation Biology." *Journal of Animal Ecology* 63: 215–44.

Chamberlin, T. C. 1890. "The Method of Multiple Working Hypotheses." *Science* 15: 92–6.

Chen, H. L., and J. L. Koprowski. 2015. "Animal Occurrence and Space Use Change in the Landscape of Anthropogenic Noise." *Biological Conservation* 192: 315–22.

Cinelli, C., A. Forney, and J. Pearl. 2022. "A Crash Course in Good and Bad Controls." *Sociological Methods and Research*: 1–30. https://journals.sagepub.com/doi/abs/10.1177/00491241221099552.

Clark, A. E., and R. Altwegg. 2019. "Efficient Bayesian Analysis of Occupancy Models with Logit Link Functions." *Ecology and Evolution* 9: 756–68.

Clarke, K. A. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22: 341–52.

Elliott, L. P., and B. W. Brook. 2007. "Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century." *Bioscience* 57: 608–14.

Fiske, I., and R. Chandler. 2011. "Unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance." *Journal of Statistical Software* 43: 1–23.

Forster, M. R. 2000. "Key Concepts in Model Selection: Performance and Generalizability." *Journal of Mathematical Psychology* 44: 205–31.

Gelman, A., and D. B. Rubin. 1995. "Avoiding Model Selection in Bayesian Social Research." *Sociological Methodology* 25: 165–73.

Gormley, A. M., D. M. Forsyth, P. Griffioen, M. Lindeman, D. S. L. Ramsey, M. P. Scroggie, and L. Woodford. 2011. "Using Presence-Only and Presence–Absence Data to Estimate the Current and Potential Distributions of Established Invasive Species." *Journal of Applied Ecology* 48: 25–34.

Grace, J. B., and K. M. Irvine. 2020. "Scientist's Guide to Developing Explanatory Statistical Models Using Causal Analysis Principles." *Ecology* 101: e02962.

Greenland, S. 2003. "Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias." *Epidemiology* 14: 300–6.

Greenland, S., J. Pearl, and J. M. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10: 37–48.

Hossack, B. R., W. H. Lowe, and P. S. Corn. 2013. "Rapid Increases and Time-Lagged Declines in Amphibian Occupancy after Wildfire." *Conservation Biology* 27: 219–28.

Johnson, J. B., and K. S. Omland. 2004. "Model Selection in Ecology and Evolution." *Trends in Ecology & Evolution* 19: 101–8.

Karavarsamis, N. 2015. "Methods for Estimating Occupancy." *Bulletin of the Australian Mathematical Society* 92: 518–9.

Laubach, Z. M., E. J. Murray, K. L. Hoke, R. J. Safran, and W. Perng. 2021. "A biologist's Guide to Model Selection and Causal Inference." *Proceedings of the Royal Society B: Biological Sciences* 288: 20202815.

Lukacs, P. M., W. L. Thompson, W. L. Kendall, W. R. Gould, P. F. Doherty, Jr., K. P. Burnham, and D. R. Anderson. 2007. "Concerns Regarding a Call for Pluralism of Information Theory and Hypothesis Testing." *Journal of Applied Ecology* 44: 456–60.

Luque-Fernandez, M. A., M. Schomaker, D. Redondo-Sanchez, M. Jose Sanchez Perez, A. Vaidya, and M. E. Schnitzer. 2019. "Educational Note: Paradoxical Collider Effect in the Analysis of Non-communicable Disease Epidemiological Data: A Reproducible Illustration and Web Application." *International Journal of Epidemiology* 48: 640–53.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. "Estimating Site Occupancy Rates when Detection Probabilities Are Less than One." *Ecology* 83: 2248–55.

Mazerolle, M. J. 2020. "AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)." R package version 2.3-1. https://cran.r-project.org/package=AICcmodavg.

McElreath, R. 2021. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*, 2nd ed. London/Boca Raton, FL: Chapman and Hall/CRC.

Midlane, N., M. J. O'Riain, G. A. Balme, H. S. Robinson, and L. T. B. Hunter. 2014. "On Tracks: A Spoor-Based Occupancy Survey of Lion *Panthera leo* Distribution in Kafue National Park, Zambia." *Biological Conservation* 172: 101–8.

Pearl, J. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82: 669–88.

Pearl, J., M. Glymour, and N. P. Jewell. 2016. *Causal Inference in Statistics: A Primer*. Chicester: John Wiley & Sons.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Richards, S. A. 2005. "Testing Ecological Theory Using the Information-Theoretic Approach: Examples and Cautionary Results." *Ecology* 86: 2805–14.

Robins, J. M., and S. Greenland. 1986. "The Role of Model Selection in Causal Inference from Nonexperimental Data." *American Journal of Epidemiology* 123: 392–402.

Rosenbaum, P. R. 1984. "The Consquences of Adjustment for a Concomitant Variable that has been Affected by the Treatment." *Journal of the Royal Statistical Society. Series A* 147: 656–66.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6: 461–4.

Shmueli, G. 2010. "To Explain or to Predict?" *Statistical Science* 25: 298–310.

Stewart, P. 2022. "Peter-Stewart/Model_selection_in_occupancy_models: v2.0.1 (v2.0.1)." Zenodo. https://doi.org/10.5281/zenodo.7043335.

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: 44–7.

Zimbres, B., C. A. Peres, G. Penido, and R. B. Machado. 2018. "Thresholds of Riparian Forest Use by Terrestrial Mammals in a Fragmented Amazonian Deforestation Frontier." *Biodiversity and Conservation* 27: 2815–36.

SUPPORTING INFORMATION

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

How to cite this article: Stewart, Peter S., Philip A. Stephens, Russell A. Hill, Mark J. Whittingham, and Wayne Dawson. 2023. "Model Selection in Occupancy Models: Inference versus Prediction." *Ecology* 104(3): e3942. https://doi.org/10.1002/ecy.3942