1   **Running Head: Model Selection in Occupancy Models**

2   **Title: Model Selection in Occupancy Models: Inference versus Prediction**

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

4   Dawson[1].

5   [1]*Department of Biosciences, Durham University, DH1 3LE, Durham, United Kingdom*

6   [2]*Department of Anthropology, Durham University, DH1 3LE, Durham, United Kingdom*

7   [3]*School of Natural and Environmental Sciences, Newcastle University, NE1 7RU, Newcastle-*

8   *Upon-Tyne, United Kingdom*

9

10   **Open Research Statement:**

11   Code to fully reproduce our simulations and analyses is available at:

12   https://zenodo.org/badge/latestdoi/462801230

13

14

15

16

17

18

19

20

21

22

Abstract

24   Occupancy models are a vital tool for applied ecologists studying the patterns and drivers of

25   species occurrence, but their use requires a method for selecting between models with

26   different sets of occupancy and detection covariates. The information-theoretic approach,

27   which employs information criteria such as Akaike's Information Criterion (AIC) is arguably

28   the most popular approach for model selection in ecology and is often used for selecting

29   occupancy models. However, the information-theoretic approach risks selecting models

30   which produce inaccurate parameter estimates, due to a phenomenon called collider bias.

31   Using simulations, we investigated the consequences of collider bias (using an illustrative

32   example called M-bias) in the occupancy and detection processes of an occupancy model,

33   and explored the implications for model selection using AIC and a common alternative, the

34   Schwarz Criterion (or Bayesian Information Criterion, BIC). We found that when M-bias was

35   present in the occupancy process, AIC and BIC selected models which inaccurately estimated

36   the effect of the focal occupancy covariate, while simultaneously producing more accurate

37   predictions of the site-level occupancy probability. In contrast, M-bias in the detection

38   process did not impact the focal estimate; all models made accurate inferences, while the site-

39   level predictions of the AIC/BIC-best model were slightly more accurate. Our results

40   demonstrate that information criteria can be used to select occupancy covariates if the sole

41   purpose of the model is prediction, but must be treated with more caution if the purpose is to

42   understand how environmental variables affect occupancy. By contrast, detection covariates

43   can usually be selected using information criteria regardless of the model's purpose. These

44   findings illustrate the importance of distinguishing between the tasks of parameter inference

45   and prediction in ecological modelling. Furthermore, our results underline concerns about the

46   use of information criteria to compare different biological hypotheses in observational

47   studies.

48   **Keywords:**

49   Akaike's information criterion, Causal inference, Collider bias, Information-theoretic

50   approach, Model selection, Occupancy models

51   **Introduction**

52   The patterns and drivers of species occurrence are of fundamental interest to ecologists.

53   Predicting where species occur enables ecologists to tackle key problems such as

54   understanding the spread of invasive species (Gormley *et al.* 2011), assessing the efficacy of

55   protected areas (Midlane *et al.* 2014), and evaluating the extinction risk (Breiner &

56   Bergamini 2018) and recovery of populations and species (Akçakaya *et al.* 2018).

57   Understanding the drivers of occurrence is also important; interventions to mitigate the

58   factors which threaten species must be informed by the diagnosis of those factors (Caughley

59   1994). Many studies have aimed to determine how occurrence is driven by factors including

60   forest degradation (Zimbres *et al.* 2018), agricultural expansion (Semper-Pascual *et al.* 2020),

61   wildfires (Hossack *et al.* 2013), and anthropogenic noise pollution (Chen & Koprowski

62   2015).

63   A key challenge in studying the patterns and drivers of occurrence is that ecologists are often

64   constrained to the use of observational data; experimental manipulations of ecological

65   systems may be physically impossible, logistically unfeasible, or unethical. Consequently,

66   one approach is to use a model which relates observed variation in species occurrence to one

67   or more environmental covariates. The model can then be used to predict species occurrence

68   at new sites, or to examine the effect of each covariate on occurrence (Shmueli 2010).

69   Occupancy models are often used for this purpose because they deal with imperfect detection

70   (MacKenzie *et al.* 2002, 2006). They do so by modelling the probability that a species

71   occupying the site is detected, often including environmental covariates to explain variation

72   in detectability among sites (MacKenzie *et al.* 2002, 2006). Occupancy models therefore

73   contain one set of covariates for occupancy probability, and a second set for detection

74   probability; the challenge is to select suitable sets of covariates to include in the model. This

75   challenge can be framed as a problem of model selection (Robins & Greenland 1986;

76   Buckland *et al.* 1997; Forster 2000; Burnham & Anderson 2004; Johnson & Omland 2004).

77   One approach to model selection is the information-theoretic approach (Anderson *et al.* 2000;

78   Burnham & Anderson 2001, 2004; Lukacs *et al.* 2007; Burnham *et al.* 2011), which

79   compares models in terms of their relative Kullback-Leibler (KL) divergence – the relative

80   distance between each model and "full reality", in units of information entropy (Forster 2000;

81   Burnham & Anderson 2001; McElreath 2021). Information criteria, of which Akaike's

82   information criterion (AIC; Akaike 1973) is the most commonly used, estimate the relative

83   KL divergence of each model using the sample data (McElreath 2021). AIC is calculated for

84   each model by taking the in-sample deviance (a measure of how well the model fits the data),

85   and adding an overfitting penalty term proportional to the number of parameters in the model

86   (Akaike 1973; Burnham *et al.* 2011). Consequently, AIC favours parsimonious models which

87   balance underfitting and overfitting, with the aim of producing better out-of-sample

88   predictions (McElreath 2021).

89   Some proponents of the information-theoretic approach argue that each model in the

90   candidate set should represent a different biological hypothesis, and that the models' relative

91   AIC scores indicate the strength of evidence for each hypothesis (Burnham *et al.* 2011).

92   However, insights from the field of causal inference reveal a potential problem with this

93   approach – collider bias. Collider bias is a form of confounding, where the estimate of an

94   effect is biased due to some feature of the data-generating process (Greenland 2003;

95   Greenland & Mansournia 2015). However, unlike the classical notion of confounding,

96   collider bias arises not from omitting an important variable from the statistical model

97 (omitted variable bias; Clarke 2005), but from including a variable that leads to bias

98 (sometimes referred to as a "bad control"; Cinelli *et al.* 2020; McElreath 2021). In the

99 simplest case, collider bias arises from including a variable which is a function of both the

100 explanatory and response variable of interest (Greenland 2003). For example, in spotted

101 hyenas (*Crocuta crocuta*), reproductive state is likely a function of both social connectedness

102 and immune function: therefore, including reproductive state as a covariate will induce a

103 spurious association between social connectedness and immune function (Laubach *et al.*

104 2021). In more complex examples with more than three variables, collider bias can arise from

105 the relationships between variables without the response variable directly affecting any other

106 variable in the system (Fig. 1).

107 As AIC and other information criteria select models based on their expected predictive

108 performance, they will not necessarily select models which return accurate parameter

109 estimates (McElreath 2021). For example, Luque-Fernandez *et al.* (2019) demonstrated that

110 for a simple collider example, including the collider covariate improved the model's AIC

111 score, while simultaneously resulting in an estimated effect size which was far from the true

112 value. It is straightforward to extend their simulation to show that this result is consistent

113 across a broad range of effect sizes (Fig. 2). However, the implications for model selection in

114 occupancy modelling, in which covariates must be selected for both the occupancy and

115 detection sub-models, have yet to be investigated.

116 To address this topic, we investigated the consequences of a form of collider bias known as

117 "M-bias" (Fig. 3 ; Greenland 2003; Cinelli *et al.* 2020) in an occupancy modelling

118 framework, and explored the implications for model selection using the information-theoretic

119 approach (using AIC). We also examined the performance of a common alternative to AIC,

120 the Schwarz criterion (or Bayesian Information Criterion, BIC; Schwarz 1978), which some

121 authors suggest can be used for selecting the "true" model from the candidate set (Aho *et al.*

122   2014). We focused on M-bias because it is a common illustrative example in the causal

123   inference literature (*e.g.* Greenland 2003; Cinelli *et al.* 2020), which demonstrates several

124   important points: i) conditioning on a variable may cause more harm than omitting it from the

125   model (Greenland *et al.* 1999), ii) collider bias can occur even when none of the variables

126   considered are directly affected by the response variable, and iii) the importance of

127   considering potential latent (unobserved) variables when drawing inferences. In our

128   simulation-based approach, we generated datasets where M-bias was present in the

129   occupancy process, the detection process, or both. We then fitted occupancy models with

130   different sets of covariates to these datasets, and evaluated them on the accuracy of their

131   parameter inferences, the accuracy of their site-level occupancy predictions, and their level of

132   support from AIC and BIC.

133

Methods

Causal inference, directed acyclic graphs and M-bias: a primer

136   Causal inference is concerned with predicting the consequences of intervening in a system, as

137   well as inferring counterfactual outcomes – events which might have happened, under

138   hypothetical unrealised conditions (Pearl 2009; Pearl *et al.* 2016). Importantly, causal

139   inference is not about 'inferring causation from correlation' – conclusions about causality

140   cannot be made from the data alone, but require causal assumptions about the process which

141   generated the data (Pearl 2009; Pearl *et al.* 2016). These assumptions, rather than remaining

142   implicit, should be communicated clearly so that they are open to scrutiny, debate, sensitivity

143   analysis, and verification (Pearl 2009).

144   Causal assumptions are typically expressed in a graphical model, usually a directed acyclic

145   graph (DAG) where the nodes represent variables and the edges represent the relationships

146   between them (Pearl 1995). The nodes on a DAG can represent both observed and latent

147   variables – the latter are usually shown surrounded by a circle (*e.g.*, A and C in Fig. 3a), and

148   we show them surrounded by brackets in the main text. The edges linking each node are

149   directional, and are thus represented as arrows; these arrows represent the mechanistic links

150   between the variables (Greenland *et al.* 1999). The succession of edges linking one variable

151   to another, regardless of which direction these edges are pointing in, is known as a path (Pearl

152   1995). For instance, Fig. 3a has two paths connecting X and ψ: $X \rightarrow \psi$, and $X \leftarrow (A) \rightarrow D \leftarrow$

153     $(C) \rightarrow \psi$.

154   A DAG corresponds to a non-parametric structural equations model (Pearl 1995), where each

155   node is represented as a function of the variables which have edges going into it, as well as a

156   "disturbance term" (ε) which represents the effect of omitted variables which are assumed to

157   be independent of one another (Pearl 1995). For example, the DAGs in Fig. 3c correspond to

158   the following set of equations:

159     $A = f_A(\varepsilon_A), \quad C = f_C(\varepsilon_C), \quad D = f_D(A, C, \varepsilon_D), \quad X = f_X(A, \varepsilon_X), \quad \psi = f_\psi(X, \varepsilon_\psi),$

160     $Q = f_Q(\varepsilon_Q), \quad V = f_V(\varepsilon_V), \quad R = f_R(Q, V, \varepsilon_R), \quad U = f_U(A, \varepsilon_U), \quad P = f_P(U, \varepsilon_P)$

161   Importantly, the model is non-parametric in the sense that the form of these functions does

162   not have to be specified (Pearl 1995; Greenland *et al.* 1999), allowing DAGs to accommodate

163   non-linear functions and interactions. For example, the variable D in the above equation

164   could be modelled as a nonlinear function of A and C, or a linear interaction between A and

165   C.

166   Once a DAG has been specified, it is analysed to identify the set of variables necessary to

167   remove structural confounds from the effect of interest. One strategy is to condition on (*i.e.*,

168   include as covariates) the variables which satisfy the "back-door criterion", in which the aim

169   is to "close all back-door paths" linking the explanatory and response variable of interest

170 (Pearl 1995). A back-door path is defined as any path which has an arrow entering the focal

171 explanatory variable (Pearl 1995) – for example, in Fig. 3a the path $X \leftarrow (A) \rightarrow D \leftarrow (C) \rightarrow$

172 ψ is a back-door path.

173 By default, paths which are a "fork" (*e.g.*, $X \leftarrow Z \rightarrow Y$) or "pipe" (*e.g.*, $X \rightarrow Z \rightarrow Y$) are

174 open, and conditioning on the middle variable (Z) closes them (Greenland 2003; Pearl *et al.*

175 2016; McElreath 2021). Paths which are a collider (*e.g.*, $X \rightarrow Z \leftarrow Y$) are closed by default,

176 and conditioning on the middle variable (Z) opens the path (Greenland *et al.* 1999; Greenland

177 2003; Pearl *et al.* 2016; McElreath 2021). A back-door path only needs to be closed in one

178 place; for instance, $X \leftarrow (A) \rightarrow D \leftarrow (C) \rightarrow \psi$ is closed because it is blocked by the collider

179 at D, irrespective of the forks at A and C.

180 The path structure of a given DAG implies that specific pairs of variables will be independent

181 of one another, conditional on a set of other variables which close the paths between them.

182 Examining whether these conditional independences hold in a given dataset is an important

183 aspect of analysing whether a DAG is consistent with the observed data (Textor *et al.* 2016).

184 The DAG displayed for ψ in Fig. 3a, for example, contains an "M"-shaped structure, which

185 can induce a type of confounding known as M-bias (Greenland 2003). By conditioning on the

186 collider (D in Fig. 3a), the back-door path is opened and the effect of interest (X on ψ in Fig.

187 3a) is confounded. When the back-door path contains latent variables (A and C in Fig. 3a) it

188 is not possible to condition on them to close the path again because they are unobserved by

189 definition. If the objective is to obtain an unconfounded estimate of the effect of X on ψ, then

190 the correct approach is to avoid conditioning on D.

191

192 **Simulation study**

193   To explore the effects of M-bias in both the occupancy and detection components of an

194   occupancy model, we conducted simulations of three different scenarios (Fig. 3) in which the

195   target for inference was the effect of variable X on occupancy probability (ψ). In the first

196   scenario (Fig. 3a), ψ was part of an M-graph while the detection probability (p) was fixed at

197   0.5. In the second scenario (Fig. 3b), ψ depended only on X, and p was now part of an M-

198   graph. In the final scenario (Fig. 3c), both the occupancy and detection probabilities were part

199   of M-graphs.

200   All three simulations followed the same process: **1)** generate a dataset with known parameter

201   values, using the relationships between variables embodied in the relevant DAGs (Fig. 3); **2)**

202   fit a number of occupancy models to the dataset (Table 1); **3)** evaluate each model's accuracy

203   in parameter estimation and prediction; **4)** evaluate each model's quality under the

204   information-theoretic framework. Each simulation was repeated 1000 times. We conducted

205   our simulations in R (v 4.0.5; R Core Team, 2021), and provide code to reproduce our

206   simulations and analyses at: https://zenodo.org/badge/latestdoi/462801230

207

208   *1) Generating a dataset*

209   Data were simulated for 3000 sites with 40 surveys each. These values were deliberately high

210   to ensure that any inaccuracy was not primarily driven by an underpowered design.

211   We first drew effect sizes for each arrow in the relevant DAG from a uniform distribution

212   between -1 and 1. Values for explanatory variables with no ingoing arrows on the DAG were

213   then drawn from a normal distribution with a mean of 0 and standard deviation of 1. We then

214   generated values for the other explanatory variables from the appropriate variables and effect

215   sizes (*i.e.*, those from ingoing arrows on the DAG; Fig. 3), plus a "disturbance term" (*sensu*

216   Pearl 1995) drawn from a normal distribution with a mean of zero and standard deviation of

217   0.025. The disturbance term was required to be non-zero to avoid perfect collinearity between

218   the explanatory variables which would result in model failure.

219   We then generated the true occupancy and detection probabilities. The log-odds of occupancy

220   and detection were first calculated as a linear function of the explanatory variables with

221   ingoing arrows on the DAG (Fig. 3), and the inverse-logit was then taken to obtain the

222   probability. The true occupancy state of each site was then simulated as a Bernoulli trial with

223   probability of success equal to the occupancy probability. We assumed that the true

224   occupancy state remained constant for each site, which is a key assumption of single-season

225   occupancy models (MacKenzie *et al.* 2002).

226   Detection histories were then generated for each site, as 40 Bernoulli trials with probability of

227   success equal to the true occupancy state multiplied by the detection probability. This means

228   that the species was never observed at sites in which it was absent (occupancy state $= 0$),

229   while when the species was present (occupancy state $= 1$) it had a chance of being observed

230   on each visit, with probability equal to the detection probability. The detection histories for

231   each site represent the observed data that would be collected by an ecologist in the field,

232   alongside the covariate values for the non-latent variables.

233

234   *2) Fitting models*

235   Occupancy models were fitted to each dataset using the *occu* function in the R package

236   *unmarked* (v.1.0.0; Fiske & Chandler, 2011), which implements the single-season occupancy

237   model developed by MacKenzie and colleagues (2002). The models used the logit link

238   function because we used the inverse-logit to simulate the data. We fitted models with

239   various combinations of observed variables (*i.e.*, excluding latent variables) for each scenario

240   (Table 1). All models included the focal covariate for occupancy (X). In scenario 1, detection

241   probability was modelled as an intercept only, as it did not depend on any covariates.

242

243   ***3) Evaluating Model Performance***

244   In each scenario, all models were evaluated on the accuracy of their parameter inferences and

245   predictions. To quantify how accurately each model estimated the effect of covariate X on the

246   occupancy probability ψ, we calculated the bias and absolute bias:

247 $$Bias = \hat{\beta}_{X\psi} - \beta_{X\psi}$$

248 $$Absolute\ Bias = |\hat{\beta}_{X\psi} - \beta_{X\psi}|$$

249   Where $\hat{\beta}_{X\psi}$ and $\beta_{X\psi}$ are the estimated and true effects of X on ψ, respectively. Additionally,

250   we constructed 95% confidence intervals for $\hat{\beta}_{X\psi}$ as follows:

251 $$95\%\ CI = \hat{\beta}_{X\psi} \pm 1.96 \times SE$$

252   where SE is the standard error of the $\hat{\beta}_{X\psi}$ estimate, obtained from the model summary table.

253   We then checked whether the true value, $\beta_{X\psi}$, was found within this interval. We also

254   checked whether the sign (positive or negative) of $\hat{\beta}_{X\psi}$ was the same as that of $\beta_{X\psi}$.

255   To evaluate each model's predictions, we used the *predict* function in R to make predictions

256   for the occupancy probability value for each site in two datasets. We first made predictions

257   for the data to which the model was fitted, to examine how the model retrodicted the sample.

258   We then examined the model's performance in out-of-sample prediction by making

259   predictions for a new dataset (also 3000 sites with 40 surveys each), which was generated

260   using the same true parameter values as the original dataset. This corresponds to the common

261   task in which occupancy models are fitted to one set of data (*e.g.*, grid squares sampled from

262   a region) and used to extrapolate to a wider set of data which are assumed to arise from the

263   same data-generating process (*e.g.*, other grid squares in the region).

264   To assess the accuracy of the model retrodictions and predictions, we first calculated the error

265   for each site:

266
$$Error = \hat{\psi}_i - \psi_i$$

267   where $\hat{\psi}_i$ and ψ_i are the estimated and true occupancy probabilities for site i, respectively. To

268   summarise this error over all sites in the dataset, we calculated the mean absolute error:

269
$$Mean\ absolute\ error = \frac{1}{n}\sum_{i=1}^{n}|\hat{\psi}_i - \psi|$$

270   where n is the number of sites. Additionally, we checked whether the true occupancy

271   probability for each site was within the 95% confidence interval obtained from the prediction,

272   and calculated the proportion of sites for which this was the case.

273

274   ***4) Evaluating models under the information-theoretic framework***

275   To examine the degree of support for each model under the information-theoretic framework

276   we obtained the AIC value for each model *m* from the model's summary table. We then

277   calculated the ΔAIC value for each model in the scenario as:

278
$$\Delta AIC_m = AIC_m - minAIC$$

279   where *AIC$_m$* is the AIC value for model *m*, and *minAIC* is the lowest AIC value for the set of

280   models in the scenario.

281   Proponents of the information-theoretic approach have advocated for multimodel inference

282   (*e.g.*, Anderson *et al.* 2000; Burnham & Anderson 2004; Burnham *et al.* 2011), in which

283   inferences are made using the entire candidate set of models, each weighted using Akaike

284   weights derived from AIC. This approach accounts for uncertainty in the model selection

285   process (Burnham & Anderson 2004). We calculated Akaike weights (w) for each model m

286   as:

287
$$w_m = \frac{\exp\left(-0.5 \times \Delta AIC_i\right)}{\sum_{r=1}^{R} \exp\left(-0.5 \times \Delta AIC_r\right)}$$

288   where R is the number of models in each scenario.

289   We also considered the Schwarz criterion (BIC; Schwarz 1978) as an alternative to AIC. BIC

290   is defined as:

291
$$BIC_m = -2\ln\left(\hat{L}\right) + j\ln(n)$$

292   where $\hat{L}$ is the maximum likelihood estimate and thus $-2\ln\left(\hat{L}\right)$ the in-sample deviance, j is

293   the number of parameters, and n is the number of observations in the dataset. We calculated

294   BIC, ΔBIC, and BIC weight values for each model using the R package *AICcmodavg*

295   (Mazerolle, 2020). ΔBIC and BIC weights are defined as for ΔAIC and AIC weights, but

296   using the BIC values for each model.

297

298   **Results**

299   *Scenario 1: M-Bias in the Occupancy Process*

300   When M-bias was present in the occupancy process, model 1 ($\psi \sim X$) estimated the true effect

301   of X on ψ much more accurately than model 2 ($\psi \sim X + D$) (Table 2, Fig. 4a,b). However,

302   comparing the two models' predictive accuracy showed the opposite picture; model 1

303   generally produced worse predictions than model 2 (Table 2, Fig. 5a,b). The models'

304   retrodictive accuracy was also qualitatively similar to their predictive accuracy (Fig. S1). AIC

305   and BIC both showed clear support for model 2 in the majority of simulations (Fig. 6a,b); in

306   85.1% of simulations model 2 received an Akaike weight of >0.99, and in 63.0% of

307   simulations it received the entire weight (Fig. 6b). The few simulations in which model 1

308   received more Akaike weight were mostly those in which $\beta_{C\psi}$ was small (Fig. S2). A similar

309   pattern of results was observed for BIC (Fig. S2), although when BIC assigned weight to

310   model 1 it generally assigned more weight than AIC (Fig. 6a).

311

Scenario 2: M-Bias in the Detection Process

313   When M-Bias was present in the detection process, both models 1 ($\psi \sim X$, $p \sim U$) and 2 ($\psi \sim$

314   X, $p \sim U + R$) accurately estimated the effect of X on ψ (Table 2, Fig. 4c,d). Both models also

315   made very accurate predictions (Table 2, Fig. 5c,d), although model 2 performed slightly

316   better - in 86.4% of simulations, the true occupancy probability was contained within model

317   2's 95% confidence interval for every predicted site, whereas model 1 accomplished this in

318   85.8% of simulations. Similar results were observed for the models' retrodictive accuracy

319   (Fig. S1c,d). Both AIC and BIC assigned the majority of weight to model 2 in most

320   simulations (Fig. 6c,d).

321

Scenario 3: M-Bias in the Occupancy and Detection Processes

323   When M-Bias was present in both the occupancy and detection processes, models 1 ($\psi \sim X$, p

324    $\sim U$) and 3 ($\psi \sim X$, $P \sim U + R$) estimated the effect of X on ψ much more accurately than

325   models 2 ($\psi \sim X + D$, $P \sim U$) and 4 ($\psi \sim X + D$, $P \sim U + R$) (Table 2, Fig. 4e-h). In general,

326   the 95% confidence interval around the estimate in models 2 and 4 only contained the true

327   value when β_{AD} and $\beta_{C\psi}$ (and to a lesser extent β_{AX}) were relatively small (Figs. S4,5). In

328   contrast, models 2 and 4 made more accurate predictions than models 1 and 3 (Table 2, Fig.

329   5e-h), and similar results were obtained for retrodictive accuracy (Fig. S1e-h). Both AIC and

330   BIC showed clear support for model 4 in the majority of simulations (Fig. 6h); the model

331   received an Akaike weight of >0.99 in 82.2% of the simulations. While model 3 did

332   occasionally receive weight, this mostly occurred when $\beta_{C\psi}$ was small (Fig. S6) and the

333   model still never received the entire weight (Fig. 6g). BIC weights were similar to the Akaike

334   weights, although BIC assigned more weight to model 3 in some simulations (Fig.4g). Again,

335   these instances were generally when $\beta_{C\psi}$ was small (Fig. S6).

336

Discussion

338   Here, we investigated the consequences of M-bias (a specific form of collider bias) for

339   occupancy modelling, and explored the implications for model selection using AIC and BIC.

340   In our simulations, we observed that when M-bias was present in the occupancy process, AIC

341   and BIC favoured a model which produced a highly inaccurate estimate of the focal effect but

342   produced more accurate predictions and retrodictions of the site-level occupancy probability.

343   In contrast, M-bias in the detection process did not result in inaccurate estimates of the focal

344   effect – both models made accurate inferences. Both models also made accurate predictions

345   and retrodictions, although those of the AIC/BIC-best models were slightly better. The same

346   results were observed when M-bias was present in both the occupancy and detection

347   processes: the model favoured by AIC and BIC produced inaccurate inferences but more

348   accurate predictions, while models made similarly accurate inferences regardless of M-bias in

349   the detection process. These results have important implications for model selection in

350   occupancy models, as well as for how the information-theoretic approach is applied in

351   ecological modelling more generally.

352

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

When M-bias was present in the occupancy process, the model which 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)*. Such biased estimates are not informative about the biological drivers underlying the observed pattern, nor do they accurately predict the consequences of intervening in the system. To illustrate the importance of this result, imagine that the fourth iteration of the first scenario, in which M-bias is present in the occupancy process only, represents a real ecological study where the objective is to inform management to increase the occupancy of a species. In this example, model 2 ($\psi \sim X + D$) received Akaike and BIC weights of one – it is clearly the "best" model, according to these criteria. The model estimated the effect of X on ψ as -1.22 ($\pm$ 0.12 standard error) – a strong negative effect on occupancy. However, this effect is confounded; in reality, the effect of X on ψ is strongly positive (0.72). Furthermore, the estimated effect of D on ψ is also strongly negative (-1.70 $\pm$ 0.09), when D actually has zero effect on occupancy. In this example, the AIC/BIC "best" model suggests that to increase ψ, we should intervene to reduce X and D. In fact, this would reduce occupancy, as well as waste resources managing the unrelated variable D. If applied to a real conservation problem informed by occupancy models (*e.g.* Hossack *et al.* 2013; Chen & Koprowski 2015; Zimbres *et al.* 2018; Semper-Pascual *et al.* 2020), the results could be disastrous. Such poor inferences were the norm in our simulations; confounded models were only able to estimate the direction of the focal effect correctly in 65.1% of cases at best – little better than the accuracy we would expect from guessing.

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 which has an open path to ψ;

378 including D provides additional information about the variation in ψ, improving prediction.

379 However, because D is a collider, including it opens the back-door path from X to ψ, biasing

380 the estimated effect of X. From the perspective of AIC and BIC, including D results in a

381 reduced in-sample deviance which typically outweighs the penalty for adding the additional

382 variable; this reduction must be greater to outweigh BIC's larger penalty term, which is why

383 BIC was more conservative in its tendency to select confounded models in our simulations

384 (Fig. 6). This also explains why AIC and BIC tended to prefer the non-confounded model

385 (omitting D) when $\beta_{C\psi}$ was close to zero (Figs. S2, S6); the near-zero effect of C on ψ meant

386 that the path from D to ψ through C was almost blocked (the other path from D to ψ was

387 blocked by conditioning on X), and therefore D explained relatively little variation in ψ.

388 In contrast to the effects of M-bias in the occupancy process, M-bias in the detection process

389 did not affect inferences about the effect of X on ψ. Additionally, including the collider

390 variable R in the detection sub-model slightly improved the accuracy of the model's

391 predictions of the site-level occupancy probability. As for the occupancy process example,

392 these results can be explained by considering how the path structure between variables will

393 affect the change in deviance when a variable is included; as the variable R has an open path

394 to p, including R explains additional variation in the detection probability, reducing the

395 deviance and allowing the model to account better for imperfect detection when estimating

396 the occupancy probability. As the detection probability is generally regarded as a nuisance

397 parameter (Karavarsamis 2015), it is inconsequential that the effect of the other detection

398 covariate (U) will be confounded. Therefore, as long as the aim is to explain as much

399 variation in the detection probability as possible, information criteria can be used for

400 selecting detection covariates.

401 The tendency for information criteria to favour confounded models with greater predictive

402 ability is not confined to collider bias. For example, simulations by McElreath (2021) showed

403   that information criteria tend to select models which condition on the mediator (M) in a pipe

404   (*e.g.*, $X \rightarrow M \rightarrow \psi$), inducing post-treatment bias (Rosenbaum 1984). Information criteria

405   prefer models which include M because M has an open path to ψ, so including M explains

406   additional variation in ψ (McElreath, 2021). However, including M also confounds the

407   estimated effect of X – not by opening a non-causal back-door path as in the collider

408   example, but instead by closing the causal front-door path which runs through M (McElreath,

409   2021). Therefore, if our aim is to estimate the effect of X on ψ then we should omit M from

410   the model, whereas if our sole aim is to predict the value of ψ then including M would

411   improve these predictions. We also expect these results to apply in other scenarios, such as

412   case control bias (Cinelli *et al.* 2020).

413

Inference and prediction are separate tasks

415   The key point supported by our results is that inference and prediction are separate tasks

416   which should not be conflated in model selection (Shmueli 2010; Laubach *et al.* 2021;

417   McElreath 2021). We echo Gelman and Rubin's (1995) criticism of selecting "a model that is

418   adequate for specific purposes without consideration of those purposes". In the context of

419   occupancy models, both explanation and prediction are important objectives, and conflating

420   the two does justice to neither. Furthermore, our results emphasise the importance of

421   considering not only the model's purpose, but also the purpose of sub-models within the

422   model; the purpose of the occupancy sub-model depends on whether we are interested in

423   predicting the occupancy state or inferring its drivers, while the detection sub-model's

424   purpose is almost always prediction of the detection probability. Consequently, how

425   occupancy covariates are chosen depends on the purpose of the model – information criteria

426   are suitable if the purpose of the model is prediction, but are unlikely to be if the purpose is

427   parameter inference – while detection covariates can generally be selected using information

428   criteria. This advice also applies to other predictive model selection methods such as cross-

429   validation; the choice of model made by AIC is asymptotically equivalent to that made by

430   leave-one-out cross validation (Stone 1977).

431

432   *Using information to compare biological hypotheses in observational studies is risky*

433   The importance of distinguishing between inference and prediction has wider implications for

434   how information-theoretic model selection is applied in ecology. Proponents of the

435   information-theoretic approach have argued that it can be used to compare multiple *a priori*

436   specified models, each representing a different biological hypothesis, with the relative AIC

437   scores indicating the strength of evidence for each hypothesis (Johnson & Omland 2004;

438   Richards 2005; Burnham *et al.* 2011). However, using information criteria in this way

439   conflates inference and prediction; information criteria select models which make better

440   predictions, but these same models can contain spurious effect sizes which hold no biological

441   meaning, while the effects of biologically important covariates are confounded. This is not

442   only the case for occupancy models; the occupancy models we employed are just an

443   extension of logistic regression (Clark & Altwegg 2019), and these points apply to other

444   forms of linear model as well (Luque-Fernandez *et al.* 2019; McElreath 2021). The

445   implication is that using information-theoretic model selection to compare biological

446   hypotheses in observational studies carries substantial risks.

447   Guarding against the risks associated with using information criteria to compare biological

448   hypotheses requires the use of subject expertise to make assumptions about the potential

449   relationships between variables in the system, to determine if phenomena such as collider

450   bias or post-treatment bias may be at play. These assumptions should be communicated

451   clearly (*e.g.*, as a DAG) so that they are open to critique and verification, and so that the

452   implications of different assumptions can be compared (Pearl 2009). Additionally, we argue

453   that if it is possible to make these assumptions, then methods such as the back-door criterion

454   can be employed to choose a suitable set of covariates for the model to estimate the effect or

455   effects of interest, circumventing the need for information criteria in these cases. Therefore,

456   we argue that if the aim is to compare biological hypotheses to infer the drivers of species

457   occurrence or to understand the consequences of management interventions, then causal

458   inference is the more suitable approach.

459

460   ***The information-theoretic approach and causal inference are complementary***

461   While we argue that comparing biological hypotheses using the information-theoretic

462   approach is risky, and that we prefer a causal inference-based approach for this purpose, we

463   must emphasise that we are not arguing that the information-theoretic approach is flawed or

464   useless for model selection. Information criteria select models from the "predictive point of

465   view" (Akaike 1998), while causal inference is concerned with estimating the effects of

466   covariates, so we see the two approaches as complementary. In the case of occupancy models

467   the two approaches may be used side-by-side in a single analysis, where occupancy

468   covariates are chosen based on causal assumptions embodied in a DAG, while the detection

469   covariates are selected using the information-theoretic approach.

470   We also argue that causal inference and the information-theoretic approach are

471   complementary because they share many philosophical underpinnings. Proponents of the

472   information-theoretic approach emphasise the importance of subject expertise and *a priori*

473   "hard thinking" to develop hypotheses which are then compared as models (Lukacs *et al.*

474   2007; Burnham *et al.* 2011). In causal inference, subject expertise and *a priori* thought are

475   vital in making the causal assumptions which are embodied in the DAG (Pearl 1995;

476   Greenland *et al.* 1999). Causal inference therefore provides a framework to support the "hard

477   thinking" required in ecological modelling (Grace & Irvine 2020). Proponents of the

478   information-theoretic approach also recognise that "a proper analysis must consider the

479   science context and cannot successfully be based on 'just the numbers'" (Burnham &

480   Anderson 2004). Similarly, proponents of causal inference argue that conclusions cannot be

481   drawn from the data alone, but require causal assumptions which come from the scientific

482   context of the model (Pearl 2009; Pearl *et al.* 2016).

483   Chamberlin's (1890) method of multiple working hypotheses is often emphasised in the

484   information-theoretic approach (Burnham & Anderson 2004; Elliott & Brook 2007). We

485   argue that causal inference is very compatible with Chamberlin's method; constructing a

486   causal model forces us to consider multiple explanations for a phenomenon, guarding against

487   the threat of "parental affection for a favourite theory" which concerned Chamberlin.

488   Furthermore, the causal approach, like the method of multiple working hypotheses, does not

489   rely on the comparison of a model with a null model or hypothesis which is often false on *a*

490   *priori* grounds and therefore provides little insight when it is rejected (Anderson *et al.* 2000;

491   Burnham & Anderson 2004). Due to the relatively static nature of causal models, we argue

492   they are especially suited to the case of multiple working hypotheses in parallel (Elliott &

493   Brook 2007), in which causation operates through multiple factors simultaneously. Moreover,

494   the tools of causal inference allow this parallel case to be extended to more complex

495   situations with indirect effects, rather than constraining our thinking to simple additive terms

496   and interactions.

497   Finally, proponents of the information-theoretic approach often emphasise George Box's

498   classic point that "all models are wrong, but some of them are useful" – they do not believe in

499   true models (Burnham & Anderson 2004). We agree, and argue that causal inference does not

500   involve making models that are true, but instead involves making models which are useful in

501   the sense of inferring the drivers of patterns and predicting the effect of interventions. Causal

502   inference provides the tools to express the assumptions underpinning these inferences in a

503   transparent manner, and allows us to examine how robust our inferences are by using

504   sensitivity analyses to explore the consequences of different causal assumptions and latent

505   confounding variables.

506

A caveat: model selection is more than selecting covariates

508   A caveat which must be borne in mind is that our article focuses on the choice of variables to

509   condition on (i.e., include as covariates), which is a key aspect of model selection, but

510   another vital part of model selection is selecting specific mathematical functions to relate

511   these variables to one-another (Johnson & Omland, 2004). However, as the rules of causal

512   inference are non-parametric in the sense that they do not assume any specific form in the

513   functions relating variable (Pearl 1995; Greenland *et al.* 1999) our conclusions hold

514   irrespective of what functional forms are chosen, and we consider any role of information

515   criteria in selecting these functions to be beyond the scope of our article.

516

Summary

518   We have demonstrated that when a form of collider bias known as M-bias is present in the

519   occupancy process, occupancy models which are favoured by AIC and BIC produce

520   inaccurate parameter estimates but accurate predictions. In contrast, M-bias in the detection

521   process does not affect the accuracy of parameter estimates. The key conclusion supported by

522   these results is that inference and prediction are separate tasks which should not be conflated

523   during model selection. The correct choice of model selection procedure depends on the

524   purpose for which the occupancy model will be used. Information-theoretic approaches are

525   suitable for selecting occupancy covariates if the model is to be used for predicting the site-

526   level occupancy probability. However, if the goal is instead to infer the effect of

527   environmental covariates on occupancy, then the use of information criteria carries

528   significant risks; we advocate for an approach based on causal inference in this situation. Our

529   results support the selection of detection covariates using information-theoretic methods

530   regardless of the model's purpose, because detection probability is almost always a nuisance

531   parameter. As the single-season occupancy models we used are in essence a form of logistic

532   regression, our results have wider implications for the use of information-theoretic model

533   selection in ecology. In particular, we argue that our results, alongside those of others

534   (Luque-Fernandez *et al.* 2019; McElreath 2021), underscore the risks associated with using

535   the information-theoretic approach to compare biological hypotheses in observational studies.

536   Causal inference and the information-theoretic approach share similar philosophical

537   underpinnings, and should be seen as complementary tools that accomplish different tasks.

538

544

545

546

547

References

549 Aho, K., Derryberry, D. & Peterson, T. (2014). Model selection for ecologists: the

550   worldviews of AIC and BIC. *Ecology*, **95**, 631–636.

551 Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood*

552   *Principle*. In: Proceeding of the Second International Symposium on Information

553   Theory (eds. Petrov, B.N. & Kaski, F.). Akademiai Kiado, Budapest, pp. 267–281.

554 Akaike, H. (1998). *Prediction and Entropy*. In: Selected Papers of Hirotugu Akaike, Springer

555   Series in Statistics (eds. Parzen, E., Tanabe, K. & Kitagawa, G.). Springer, New York,

556   NY, pp. 387–410.

557 Akçakaya, H.R., Bennett, E.L., Brooks, T.M., Grace, M.K., Heath, A., Hedges, S., et al.

558   (2018). Quantifying species recovery and conservation success to develop an IUCN

559   Green List of Species. *Conservation Biology*, **32**, 1128–1138.

560 Anderson, D.R., Burnham, K.P. & Thompson, W.L. (2000). Null Hypothesis Testing:

561   Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management*, **64**,

562   912–923.

563 Breiner, F.T. & Bergamini, A. (2018). Improving the estimation of area of occupancy for

564   IUCN Red List assessments by using a circular buffer approach. *Biodiversity*

565   *Conservation*, **27**, 2443–2448.

566 Buckland, S.T., Burnham, K.P. & Augustin, N.H. (1997). Model Selection: An Integral Part

567   of Inference. *Biometrics*, **53**, 603–618.

568 Burnham, K.P. & Anderson, D.R. (2001). Kullback-Leibler information as a basis for strong

569   inference in ecological studies. *Wildlife Research*., **28**, 111.

570 Burnham, K.P. & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and

571   BIC in Model Selection. *Sociological Methods & Research*, **33**, 261–304.

572   Burnham, K.P., Anderson, D.R. & Huyvaert, K.P. (2011). AIC model selection and

573   multimodel inference in behavioral ecology: some background, observations, and

574   comparisons. *Behavioral Ecology and Sociobiology,* **l**, 65, 23–35.

575   Caughley, G. (1994). Directions in Conservation Biology. Journal of Animal Ecology, 63,

576   215–244.

577   Chamberlin, T.C. (1890). The Method of Multiple Working Hypotheses. *Science*, **15**, 6.

578   Chen, H.L. & Koprowski, J.L. (2015). Animal occurrence and space use change in the

579   landscape of anthropogenic noise. *Biological Conservation*, **192**, 315–322.

580   Cinelli, C., Forney, A. & Pearl, J. (2020). A Crash Course in Good and Bad Controls. *SSRN*

581   *Journal*.

582   Clark, A.E. & Altwegg, R. (2019). Efficient Bayesian analysis of occupancy models with

583   logit link functions. *Ecology and Evolution*, **9**, 756–768.

584   Clarke, K.A. (2005). The Phantom Menace: Omitted Variable Bias in Econometric Research.

585   *Conflict Management and Peace Science*, **22**, 341–352.

586   Elliott, L.P. & Brook, B.W. (2007). Revisiting Chamberlin: Multiple Working Hypotheses

587   for the 21st Century. *BioScience*, **57**, 608–614.

588   Fiske, I., & Chandler, R. (2011). unmarked: An R Package for Fitting Hierarchical Models

589   of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, **43**, 1-23.

590   Forster, M.R. (2000). Key Concepts in Model Selection: Performance and Generalizability.

591   *Journal of Mathematical Psychology*, **44**, 205–231.

592   Gelman, A. & Rubin, D.B. (1995). Avoiding Model Selection in Bayesian Social Research.

593   *Sociological Methodology*, **25**, 165–173.

594   Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S.L., Scroggie,

595   M.P., et al. (2011). Using presence-only and presence–absence data to estimate the

596   current and potential distributions of established invasive species. *Journal of Applied*

597   *Ecology*, **48**, 25–34.

598   Grace, J.B. & Irvine, K.M. (2020). Scientist's guide to developing explanatory statistical

599   models using causal analysis principles. *Ecology*, **101**, e02962.

600   Greenland, S. (2003). Quantifying Biases in Causal Models: Classical Confounding vs

601   Collider-Stratification Bias. *Epidemiology*, **14**, 300–306.

602   Greenland, S. & Mansournia, M.A. (2015). Limitations of individual causal models, causal

603   graphs, and ignorability assumptions, as illustrated by random confounding and

604   design unfaithfulness. *European Journal of Epidemiology*, **30**, 1101–1110.

605   Greenland, S., Pearl, J. & Robins, J.M. (1999). Causal diagrams for epidemiologic research.

606   *Epidemiology*, **10**, 37–48.

607   Hossack, B.R., Lowe, W.H. & Corn, P.S. (2013). Rapid Increases and Time-Lagged Declines

608   in Amphibian Occupancy after Wildfire. *Conservation Biology*, **27**, 219–228.

609   Johnson, J.B. & Omland, K.S. (2004). Model selection in ecology and evolution. *Trends in*

610   *Ecology & Evolution*, **19**, 101–108.

611   Karavarsamis, N. (2015). Methods for estimating occupancy. *Bulletin of the Australian*

612   *Mathematical Society,* **92**, 518–519.

613   Laubach, Z.M., Murray, E.J., Hoke, K.L., Safran, R.J. & Perng, W. (2021). A biologist's

614   guide to model selection and causal inference. *Proceedings of the Royal Society B:*

615   *Biological Sciences*, **288**, 20202815.

616   Lukacs, P.M., Thompson, W.L., Kendall, W.L., Gould, W.R., Doherty, P.F., Burnham, K.P.,

617   et al. (2007). Concerns regarding a call for pluralism of information theory and

618   hypothesis testing: Concerns about pluralism. *Journal of Applied Ecology*, **44**, 456–

619   460.

620 Luque-Fernandez, M.A., Schomaker, M., Redondo-Sanchez, D., Jose Sanchez Perez, M.,

621   Vaidya, A. & Schnitzer, M.E. (2019). Educational Note: Paradoxical collider effect in

622   the analysis of non-communicable disease epidemiological data: a reproducible

623   illustration and web application. *International Journal of Epidemiology*, **48**, 640–653.

624 MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A.

625   (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less

626   Than One. *Ecology*, **83**, 2248–2255.

627 MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006).

628   *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species*

629   *Occurrence*. 1st. edn. Elsevier.

630 Mazerolle, M.J. (2020). *AICcmodavg: Model selection and multimodel inference based on*

631   *(Q)AIC(c)*. R package version 2.3-1. https://cran.r-project.org/package=AICcmodavg.

632 McElreath, R. (2021). *Statistical Rethinking: A Bayesian Course with Examples in R and*

633   *STAN*. 2nd edn. Chapman and Hall/CRC.

634 Midlane, N., O'Riain, M.J., Balme, G.A., Robinson, H.S. & Hunter, L.T.B. (2014). On

635   tracks: A spoor-based occupancy survey of lion Panthera leo distribution in Kafue

636   National Park, Zambia. *Biological Conservation*, **172**, 101–108.

637 Okes, N.C. & O'Riain, M.J. (2017). Otter occupancy in the Cape Peninsula: Estimating the

638   probability of river habitat use by Cape clawless otters, Aonyx capensis, across a

639   gradient of human influence. *Aquatic Conservation: Marine and Freshwater*

640   *Ecosystems*, **27**, 706–716.

641 Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, **82**, 669–688.

642 Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, **3**, 96–146.

643 Pearl, J., Glymour, M. & Jewell, N.P. (2016). *Causal Inference in Statistics: A Primer*. John

644   Wiley & Sons, Incorporated, Chicester, United Kingdom.

645   R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation

646   for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

647   Richards, S.A. (2005). Testing Ecological Theory Using the Information-Theoretic

648   Approach: Examples and Cautionary Results. *Ecology*, **86**, 2805–2814.

649   Robins, J.M. & Greenland, S. (1986). The role of model selection in causal inference from

650   nonexperimental data. *American Journal of Epidemiology*, **123**, 392–402.

651   Rosenbaum, P.R. (1984). The Consquences of Adjustment for a Concomitant Variable That

652   Has Been Affected by the Treatment. *Journal of the Royal Statistical Society. Series A*

653   *(General)*, **147**, 656–666.

654   Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461–

655   464.

656   Semper-Pascual, A., Decarre, J., Baumann, M., Camino, M., Di Blanco, Y., Gómez-Valencia,

657   B., et al. (2020). Using occupancy models to assess the direct and indirect impacts of

658   agricultural expansion on species' populations. *Biodiversity Conservation*, **29**, 3669–

659   3688.

660   Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, **25**, 298-310.

661   Stone, M. (1977). An Asymptotic Equivalence of Choice of Model by Cross-Validation and

662   Akaike's Criterion. *Journal of the Royal Statistical Society. Series B*

663   *(Methodological)*, **39**, 44–47.

664   Zimbres, B., Peres, C.A., Penido, G. & Machado, R.B. (2018). Thresholds of riparian forest

665   use by terrestrial mammals in a fragmented Amazonian deforestation frontier.

666   *Biodiversity Conservation*, **27**, 2815–2836.

667

668

669

670 **Tables**

671 **Table 1.** Occupancy models fitted in each simulation scenario. All models also included an

672 intercept term for both occupancy and detection. *X* represents an explanatory variable for

673 which we are interested in estimating the effect on the occupancy probability. *D*, *U*, and *R*

674 represent additional variables which are thought to be important in the occupancy or

675 detection processes.

Model	Occupancy covariates	Detection covariates
Scenario 1		
Model 1	X	n/a
Model 2	X + D	n/a
Scenario 2		
Model 1	X	U
Model 2	X	U + R
Scenario 3		
Model 1	X	U
Model 2	X + D	U
Model 3	X	U + R
Model 4	X + D	U + R

676

677

678

679

680

681

682

683

684 **Table 2.** Inferential and predictive accuracy of occupancy models across all simulations

685 (n=1000) for each of the three scenarios: M-bias in the occupancy process only (left), M-bias

686 in the detection process only (centre), and M-bias in both the occupancy and detection

687 processes (right). Each model is shown as a separate column, with the occupancy *(ψ)* and

688 detection *(p)* covariates displayed at the top. Each row indicates a separate metric

689 summarising the models' performance across the simulations. Full definitions of each metric

690 can be found in the main text.

			M-bias in occupancy process		M-bias in detection process		M-bias in occupancy and detection processes			
			$\psi \sim X$	$\psi \sim X+D$	$\psi \sim X$ $p \sim U$	$\psi \sim X$ $p \sim U+R$	$\psi \sim X$ $p \sim U$	$\psi \sim X+D$ $p \sim U$	$\psi \sim X$ $p \sim U+R$	$\psi \sim X+D$ $p \sim U+R$
Inference	Bias	Q1	-0.07	-1.13	-0.03	-0.03	-0.07	0.93	-0.07	0.93
		Mean	-0.001	0.03	<0.001	<0.001	0.04	0.19	0.05	0.19
		Q3	0.06	0.88	0.03	0.03	0.06	0.99	0.06	0.99
	Absolute bias	Q1	0.03	0.29	0.01	0.01	0.03	0.32	0.03	0.32
		Median	0.07	1.00	0.03	0.03	0.06	0.95	0.06	0.95
		Q3	0.13	2.69	0.05	0.05	0.13	2.47	0.13	2.47
	True value in 95% C.I.		89.6%	20.07%	95.4%	95.4%	91.5%	20.2%	91.6%	20.1%
	Sign correct		94.5%	65.1%	98.3%	98.3%	93.6%	63.3%	93.6%	63.3%
Prediction	Mean absolute error	Q1	0.048	0.001	0.006	0.006	0.047	0.010	0.047	0.010
		Median	0.092	0.015	0.009	0.009	0.087	0.014	0.087	0.014
		Q3	0.128	0.022	0.012	0.012	0.126	0.021	0.126	0.021
	Proportion of sites with true ψ in 95% C.I.	Q1	10.6%	65.4%	100%	100%	10.8%	69.0%	10.8%	69.0%
		Median	15.4%	89.0%	100%	100%	16.2%	90.1%	16.2%	90.2%
		Q3	29.4%	99.9%	100%	100%	30.2%	99.9%	30.1%	99.9%

691

692   **Figure Captions**

693   **Figure 1.** Directed acyclic graph showing a hypothetical example in which we are interested

694   in estimating the direct effect of the density of an invasive plant on animal occupancy (red

695   arrow). Invasive plant density also influences the density of a native food plant, which in turn

696   influences the animal's occupancy. Soil properties influence the density of both the native

697   and invasive plant – these properties are latent (unobserved), and so are shown in a circle.

698   Urbanisation influences the densities of both plant species, as well as animal occupancy –

699   here, urbanisation is also latent. In this example, native food plant density is a collider which

700   sits on the back-door path from invasive plant density to animal occupancy, *Invasive plant ←*

701   *(Soil properties) → Native food plant ← (Urbanisation) → Animal occupancy*. Including

702   native food plant density in the model opens the back-door path, biasing the estimated direct

703   effect of the invasive plant. Notably, in this example the effect of interest cannot be estimated

704   by including any of the available covariates – measuring and including urbanisation is

705   necessary.

706   **Figure 2.** Luque-Fernandez *et al.* (2019) presented simulations illustrating classical

707   confounding and collider bias in a simple linear model. We extended their example by

708   conducting 10000 simulation iterations for each example, with each iteration using true effect

709   sizes drawn from a uniform distribution between -2 and 2 to ensure that the results are not

710   specific to the choice of parameter values. In the classical confounding example (**A**),

711   including the variable Z reduces the absolute bias (the absolute difference between the

712   estimated and true effect: $|\hat{\beta}_{XY} - \beta_{XY}|$) when estimating the effect of X on Y. Conversely, in

713   the collider example (**B**), including Z increases the absolute bias. However, in both cases

714   Akaike's Information Criterion (AIC) favours the model which includes Z (**C**), illustrating

715   that AIC does not always favour models which produce accurate parameter estimates. R code

716   to reproduce the simulations is available at: https://zenodo.org/badge/latestdoi/462801230

717   **Figure 3.** Directed acyclic graphs (DAGs) showing the relationship between variables in the

718   occupancy and detection processes of the three simulation scenarios: **a)** scenario 1 (M-bias in

719   the occupancy process), **b)** scenario 2 (M-bias in the detection process), **c)** scenario 3 (M-bias

720   in both the occupancy and detection processes). ψ represents the probability that a site is

721   occupied by a species (occupancy probability), while p represents the probability that a

722   species is detected given that it is present at a site (detection probability). X represents an

723   explanatory variable for which we are interested in estimating the effect on the occupancy

724   probability. A, C, D, Q, U, R and V represent additional variables which are thought to be

725   important in the system. Circled variables are latent (unobserved).

726   **Figure 4.** True versus estimated effect of X on occupancy probability (ψ), for the following

727   occupancy models: **A)** scenario 1, model 1; **B)** scenario 1, model 2; **C)** scenario 2, model 1;

728   **D)** scenario 2, model 2; **E)** scenario 3, model 1; **F)** scenario 3, model 2; **G)** scenario 3, model

729   3; **H)** scenario 4, model 4. Each point represents the result from one simulation, with 1000

730   simulations in total. Note that the y-axis is truncated at -10 and 10; plots B, F, and H omit 40,

731   38 and 38 points respectively which lay outside this range. Blue points indicate that the true

732   value was contained within the estimate's 95% confidence interval, while unfilled circles

733   indicate that the true value was not contained within the interval. Dashed black line indicates

734   equality between the true and estimated effect – increased vertical distance from this line

735   indicates a more biased estimate. Each model's covariates for ψ and the detection probability

736   (p) are shown above their respective plot. For explanation of the scenarios, see Table 1 and

737   main text.

738   **Figure 5.** Kernel density estimate contours showing two measures of predictive accuracy

739   when predicting site-level occupancy probability (ψ), for 1000 simulations. The x-axis shows

740   the proportion of sites (out of 3000 sites) for which the true occupancy probability was

741   contained within the 95% confidence interval around the model's prediction. The y-axis

742   shows the mean absolute error of the predictions. Consequently, the bottom right of each plot

743   indicates higher predictive accuracy, while the top left indicates lower predictive accuracy.

744   The density of simulations within this area is shown by the coloured contours, with lighter

745   colours indicating a higher density of simulations. Results are displayed for the following

746   occupancy models: **A)** scenario 1, model 1; **B)** scenario 1, model 2; **C)** scenario 2, model 1;

747   **D)** scenario 2, model 2; **E)** scenario 3, model 1; **F)** scenario 3, model 2; **G)** scenario 3, model

748   3; **H)** scenario 4, model 4. Each model's covariates for ψ and the detection probability (p) are

749   shown above their respective plot. For explanation of the scenarios, see Table 1 and main

750   text.

751   **Figure 6.** Akaike weight (yellow area) and BIC weight (blue area) for 1000 simulations of

752   eight occupancy models. Simulations are shown ranked by weight, with higher Akaike and

753   BIC weights shown on the right. Higher weights correspond to lower AIC/BIC values,

754   indicating more support for the model. The panels display: **A)** scenario 1, model 1; **B)**

755   scenario 1, model 2; **C)** scenario 2, model 1; **D)** scenario 2, model 2; **E)** scenario 3, model 1;

756   **F)** scenario 3, model 2; **G)** scenario 3, model 3; **H)** scenario 4, model 4. Each model's

757   covariates for the occupancy probability (ψ) and the detection probability (p) are shown

758   above their respective plot. For explanation of the scenarios, see Table 1 and main text.

759   Dashed horizontal lines are shown for weights of 0, 0.5, and 1.

760

761

762

763

764
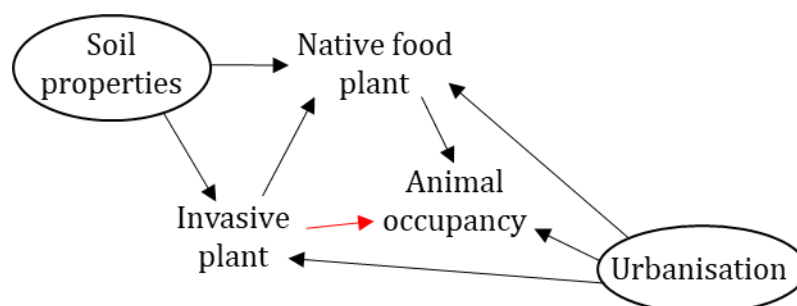
765 **Figures**

766

767 **Figure 1.** Directed acyclic graph showing a hypothetical example in which we are interested
768 in estimating the direct effect of the density of an invasive plant on animal occupancy (red
769 arrow). Invasive plant density also influences the density of a native food plant, which in turn
770 influences the animal's occupancy. Soil properties influence the density of both the native
771 and invasive plant – these properties are latent (unobserved), and so are shown in a circle.
772 Urbanisation influences the densities of both plant species, as well as animal occupancy –
773 here, urbanisation is also latent. In this example, native food plant density is a collider which
774 sits on the back-door path from invasive plant density to animal occupancy, *Invasive plant ←*
775 *(Soil properties) → Native food plant ← (Urbanisation) → Animal occupancy*. Including
776 native food plant density in the model opens the back-door path, biasing the estimated direct
777 effect of the invasive plant. Notably, in this example the effect of interest cannot be estimated
778 by including any of the available covariates – measuring and including urbanisation is
779 necessary.

780

781

782
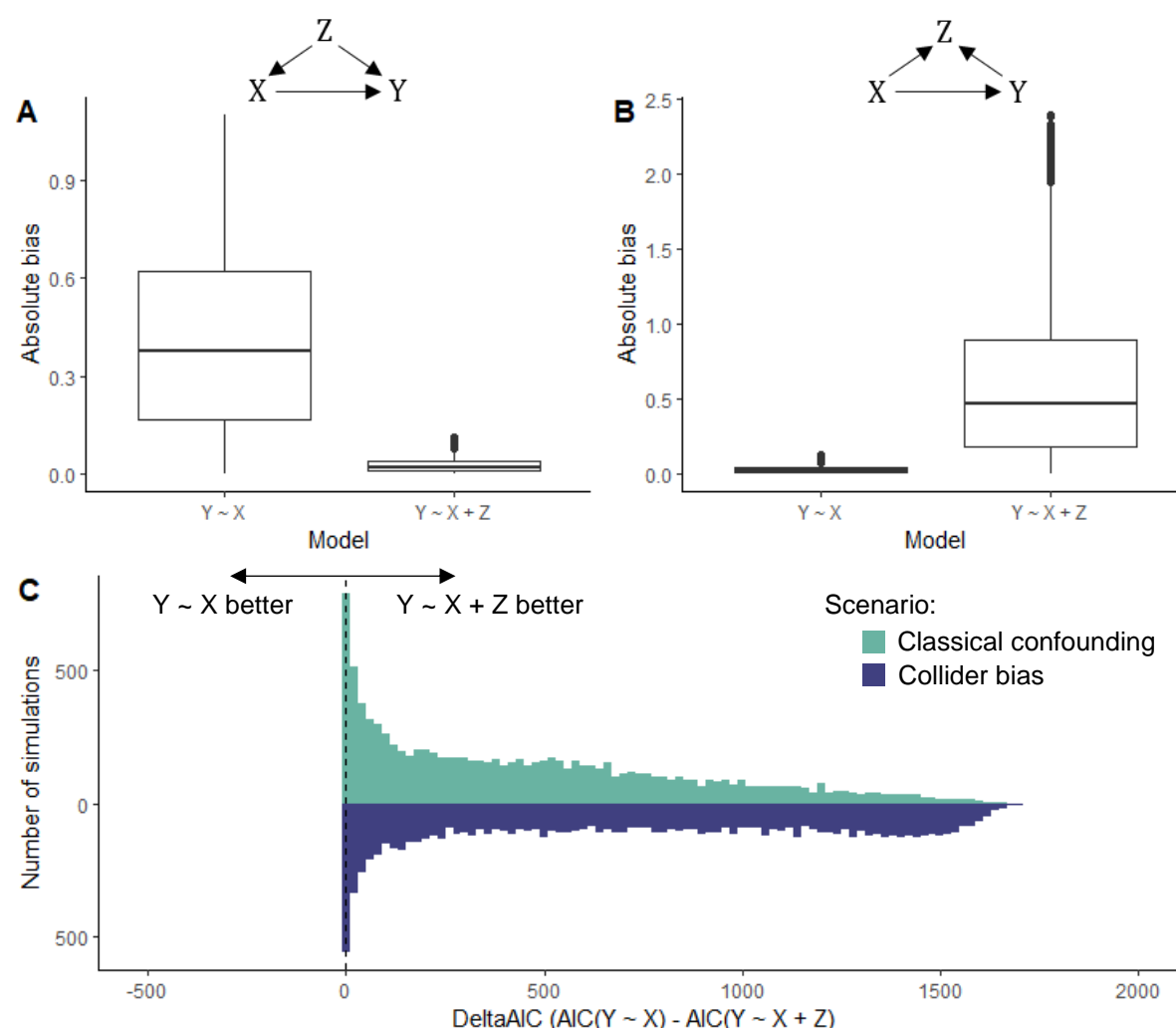
783

784

785

786

787

788

789

Figure 2. Luque-Fernandez *et al.* (2019) presented simulations illustrating classical confounding and collider bias in a simple linear model. We extended their example by conducting 10000 simulation iterations for each example, with each iteration using true effect sizes drawn from a uniform distribution between -2 and 2 to ensure that the results are not specific to the choice of parameter values. In the classical confounding example (**A**), including the variable Z reduces the absolute bias (the absolute difference between the estimated and true effect: $|\hat{\beta}_{XY} - \beta_{XY}|$) 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) favours the model which includes Z (**C**), illustrating that AIC does not always favour models which produce accurate parameter estimates. R code to reproduce the simulations is available at: https://zenodo.org/badge/latestdoi/462801230
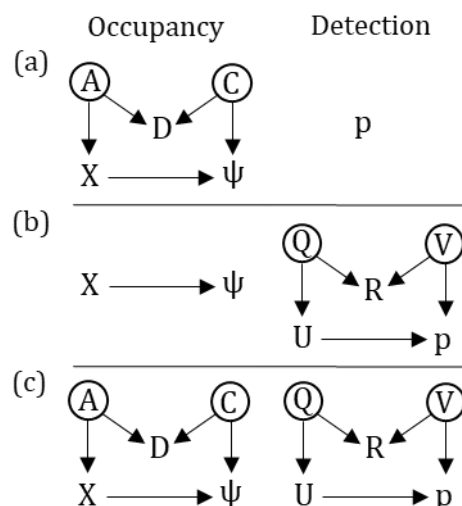
Figure 3. Directed acyclic graphs (DAGs) showing the relationship between variables in the occupancy and detection processes of the three simulation scenarios: **a)** scenario 1 (M-bias in the occupancy process), **b)** scenario 2 (M-bias in the detection process), **c)** scenario 3 (M-bias in both the occupancy and detection processes). ψ represents the probability that a site is occupied by a species (occupancy probability), while p represents the probability that a species is detected given that it is present at a site (detection probability). X represents an explanatory variable for which we are interested in estimating the effect on the occupancy probability. A, C, D, Q, U, R and V represent additional variables which are thought to be important in the system. Circled variables are latent (unobserved).

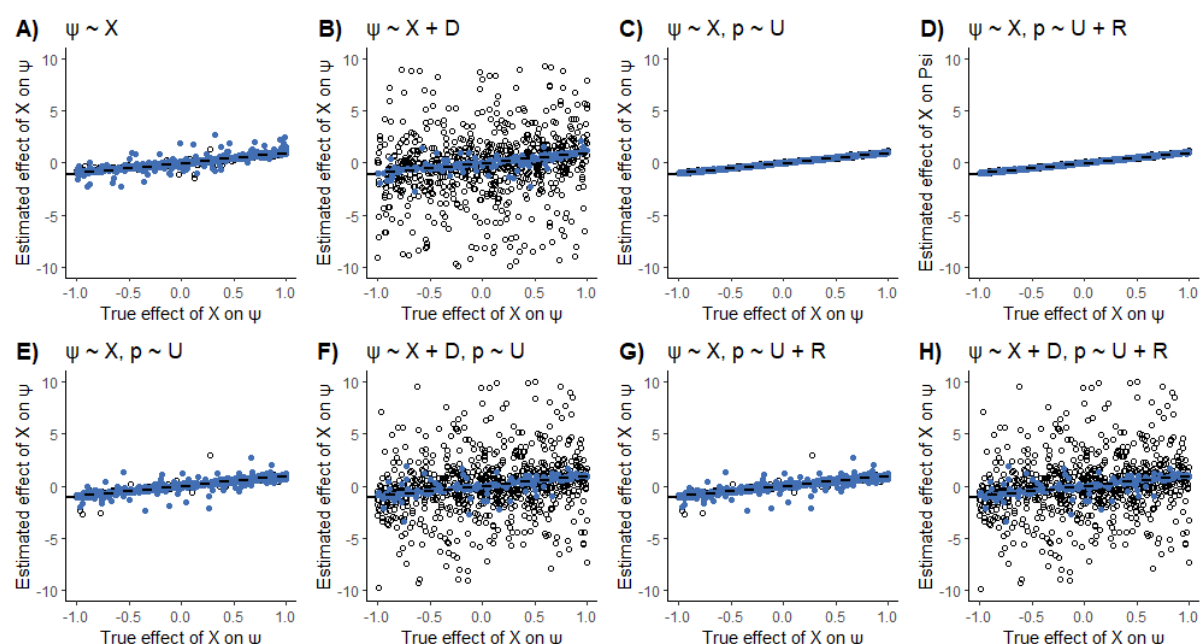**Figure 4.** True versus estimated effect of *X* on occupancy probability (ψ), for the following occupancy models: **A)** scenario 1, model 1; **B)** scenario 1, model 2; **C)** scenario 2, model 1; **D)** scenario 2, model 2; **E)** scenario 3, model 1; **F)** scenario 3, model 2; **G)** scenario 3, model 3; **H)** scenario 4, model 4. Each point represents the result from one simulation, with 1000 simulations in total. Note that the y-axis is truncated at -10 and 10; plots B, F, and H omit 40, 38 and 38 points respectively which lay outside this range. Blue points indicate that the true value was contained within the estimate's 95% confidence interval, while unfilled circles indicate that the true value was not contained within the interval. Dashed black line indicates equality between the true and estimated effect – increased vertical distance from this line indicates a more biased estimate. Each model's covariates for ψ and the detection probability (p) are shown above their respective plot. For explanation of the scenarios, see Table 1 and main text.
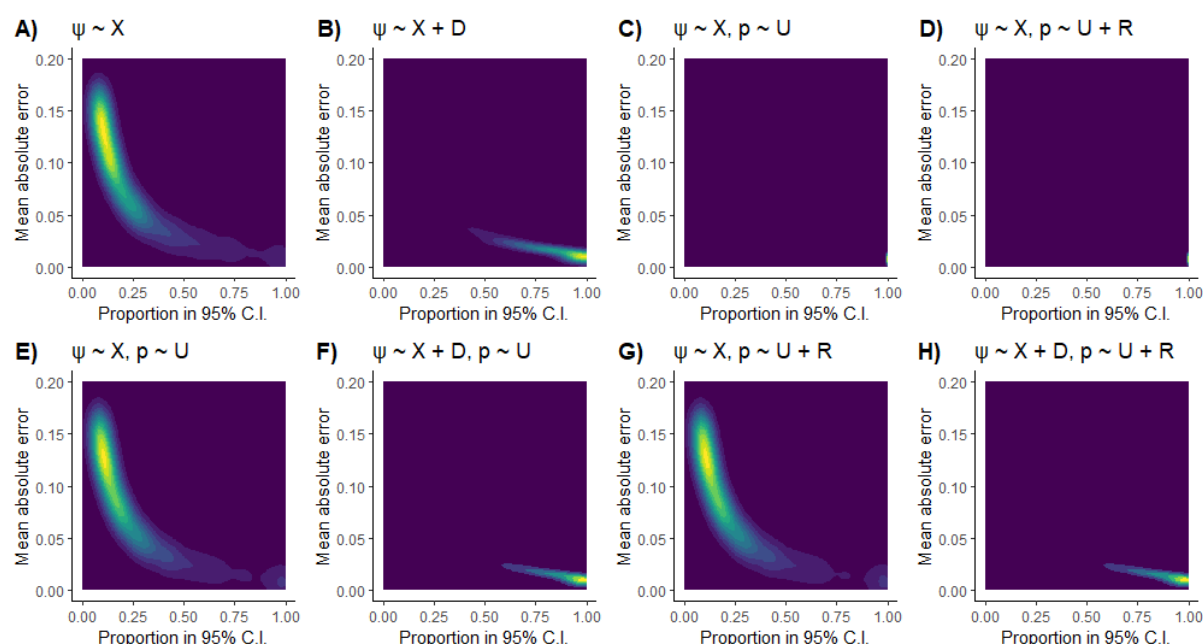
Figure 5. Kernel density estimate contours showing two measures of predictive accuracy when predicting site-level occupancy probability (ψ), for 1000 simulations. The x-axis shows the proportion of sites (out of 3000 sites) for which the true occupancy probability was contained within the 95% confidence interval around the model's prediction. The y-axis shows the mean absolute error of the predictions. Consequently, the bottom right of each plot indicates higher predictive accuracy, while the top left indicates lower predictive accuracy. The density of simulations within this area is shown by the coloured contours, with lighter colours 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. For explanation of the scenarios, see Table 1 and main text.
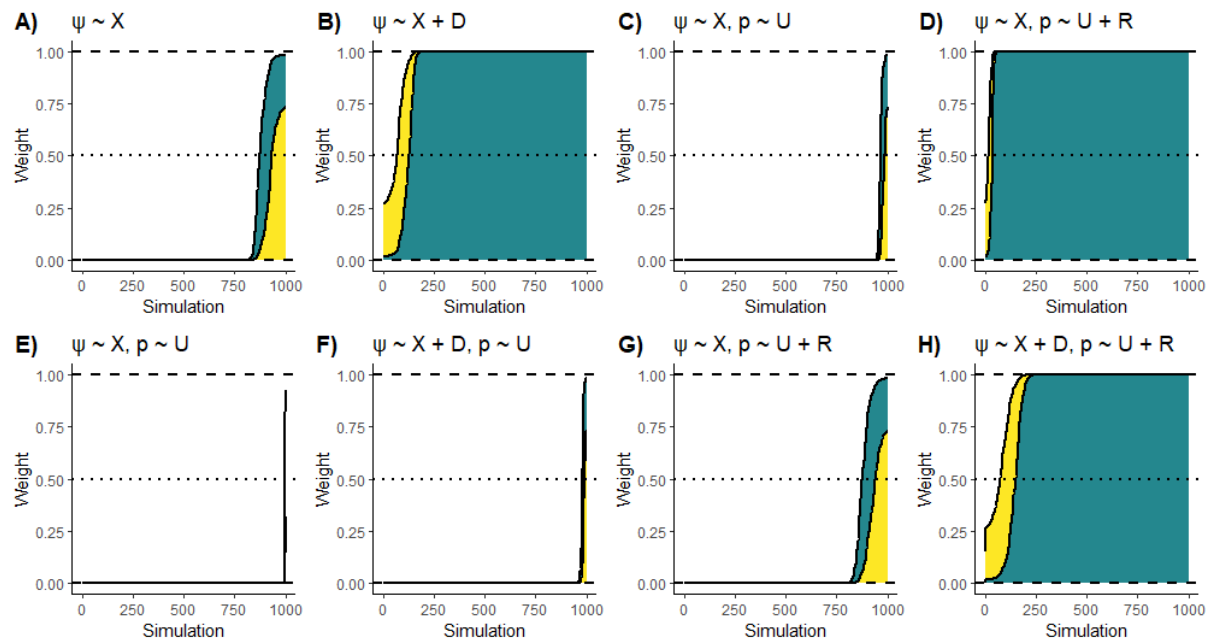
Figure 6. Akaike weight (yellow area) and BIC weight (blue area) for 1000 simulations of eight occupancy models. Simulations are shown ranked by weight, with higher Akaike and BIC weights shown on the right. Higher weights correspond to lower AIC/BIC values, indicating more support for the model. The 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. For explanation of the scenarios, see Table 1 and main text. Dashed horizontal lines are shown for weights of 0, 0.5, and 1.