

Using causal diagrams and superpopulation models to correct geographic biases in biodiversity monitoring data

*¹Robin J. Boyd, ¹Marc Botham, ²Emily Dennis, ²Richard Fox, ¹Collin Harrower, ²Ian Middlebrook, ^{1,3}David B. Roy and ¹Oliver L. Pescott

¹UK Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, UK

²Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset, UK

³Centre for Ecology and Conservation, University of Exeter, Falmouth, Cornwall, UK

*Corresponding author email: robboy@ceh.ac.uk

Running headline: Correcting sampling bias in biodiversity data

Abstract

1. Biodiversity monitoring schemes periodically measure species' abundances and distributions at a sample of sites to understand how they have changed over time. Often, the aim is to infer change in an average sense across some wider landscape. Inference to the wider landscape is simple if the species' abundances and distributions are similar at sampled to non-sampled locations. Otherwise, the data are geographically biased, and some form of correction is desirable.
2. We combine causal diagrams with "superpopulation models" to correct time-varying geographic biases in biodiversity monitoring data. For a given time-period, expert-derived causal diagrams are used to deduce the set of variables that explain the geographic bias, and superpopulation models adjust for these variables to produce a corrected estimate of a landscape-wide mean of e.g. abundance or occupancy. Estimating a time trend in the variable of interest is achieved by fitting models for multiple time-periods and, if the drivers of bias are suspect to change over time, by constructing per period causal diagrams. We test the approach using simulated data then apply it to real data from the UK Butterfly Monitoring Scheme (UKBMS).
3. Where the variables that explain the geographic bias are known and measured without error, our method is unbiased. Introducing measurement error reduces the method's efficacy, but it is still an improvement on using the sample mean. When applied to data from the UKBMS, the approach gives different results to the scheme's current method, which assumes no geographic bias.
4. Where the goal is to estimate change in some variable of interest at the landscape level (e.g. biodiversity indicators), models that do not adjust for geographic bias implicitly assume it does not exist. Our approach makes the weaker assumption that there is no geographic bias conditional on the adjustment variables, so it should yield more accurate estimates of time trends in many circumstances. The method does require assumptions about the drivers of bias, but these are codified explicitly in the causal diagrams. Operationalising our approach should be less costly than full probability sampling, which would be needed to satisfy the assumptions of conventional approaches.

Key words: Directed Acyclic Graph; expert consultation; imputation; sampling bias; species abundance; time trend

42 Introduction

43 Monitoring biodiversity

44 The goal of biodiversity monitoring is to understand how some variable of interest Y , often when
45 summarised across the relevant landscape or study area, has changed over time (cf. Yoccoz et al.,
46 2001). Y might be a species' (relative) abundance, occupancy or some summary thereof across many
47 species. The landscape is often delimited by national boundaries (Moussy et al., 2022), but in
48 principle it could be any geographic area. Y is usually summarised across the landscape periodically
49 by taking its mean or something similar (e.g. a categorical year effect from a model in which Y is the
50 response; Brereton, Roy, et al., 2011; Powney et al., 2019; Stroh et al., 2023).

51 To calculate the landscape-wide mean of Y directly, one would need to know its distribution across
52 the landscape as a whole (i.e. its value at every location). Generally, however, the landscape
53 distribution of Y is not known, because it is not possible to collect data at every location. Instead, the
54 usual strategy is to split the landscape into areal units, or “sites”, and to collect data at some of them.
55 The landscape-wide mean of Y is then *estimated* from the sample of sites for which data are available.

56 Geographic sampling biases

57 When estimating the landscape-wide mean of Y from a sample of sites, there is a risk of geographic
58 sampling bias. A dataset is geographically biased if the distribution of Y across sampled sites differs
59 from its distribution across the wider landscape. An equivalent definition is the existence of a non-
60 zero correlation between Y and a sample inclusion indicator R , which takes the value 1 at sampled
61 sites and 0 elsewhere (Boyd, Powney, et al., 2023; Meng, 2018). If the correlation is positive, then the
62 bias is positive (i.e. the mean of Y is larger across sampled sites than across the landscape as a whole),
63 and vice versa. The greater the magnitude of the correlation, the more severe the bias.

64 Biodiversity monitoring data are often collected in such a way that the same factors affect R and Y ,
65 which induces a correlation between the two (and therefore a bias). To give one simple example
66 (many others are possible), Y might be some species' abundance, which is positively affected by
67 habitat quality. Habitat quality might also have a positive effect on R , because data collectors are
68 often volunteer naturalists (i.e. citizen scientists) and prefer to visit sites that are interesting in terms
69 of wildlife (Bowler et al., 2022; Forister et al., 2023). In this situation, Y would be larger at sampled
70 sites, which is to say that there would be a positive bias. The reverse would be true if sampling was
71 more likely at sites where species are faring poorly: say, in built-up areas that have poor quality
72 habitat, but which are easy for recorders to access by road (Hughes et al., 2020).

73 Data from structured monitoring schemes are less susceptible to geographic biases than those from
74 most other sources (Geldmann et al., 2016), but they are not immune. Many structured monitoring
75 schemes aim to collect data at a random sample of sites (e.g. Pescott et al., 2019; Robbins et al.,
76 1986). If implemented properly, random sampling ensures no correlation between R and Y in
77 expectation (i.e. on average over many possible samples; see Meng, 2018). Obtaining a true random
78 sample is challenging, however, especially if there is a reliance on volunteers to collect the data
79 (which is true of many monitoring schemes). The volunteers might be unwilling or unable to visit
80 some randomly selected sites, whether because they are difficult to access, uninteresting in terms of
81 wildlife or for some other reason (Pescott et al., 2015). Nominally random samples with incomplete
82 uptake of sites, or “nonresponse” in survey sampling parlance, are almost certain to be biased (Bailey,
83 2023).

84 Mitigating geographic biases

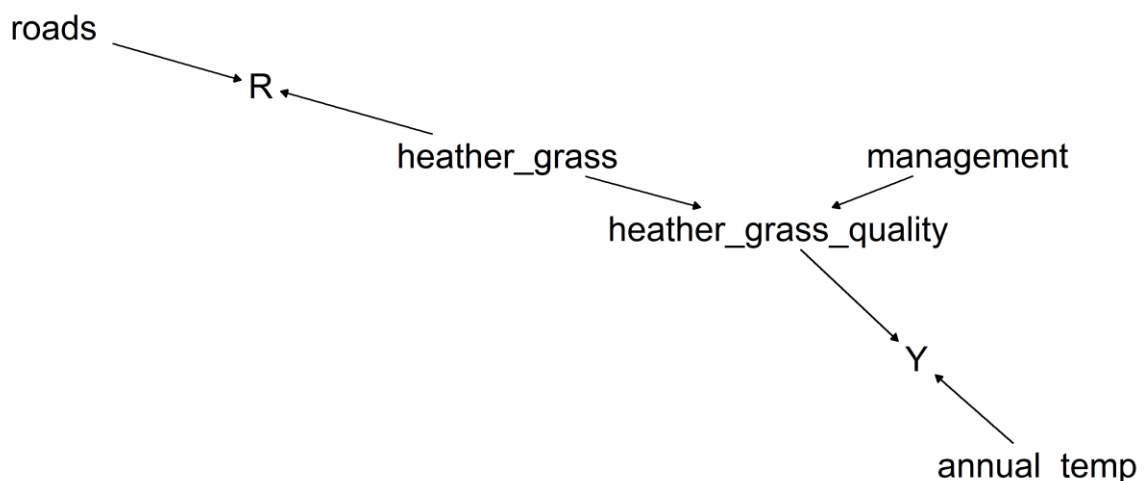
85 Various approaches exist to mitigate sampling biases, most of which exploit the fact that some set of
86 variables A could explain the correlation between R and Y (Lohr, 2022; Meng, 2022). If, to use the
87 above example, R and Y are both greater at sites with good quality habitat, then the presence of that
88 habitat will explain some of the correlation between the two. Where the variables in A can be
89 identified—a point we come back to below—and are reflected in available data, statistically
90 “adjusting for” them will reduce the sampling bias (Collins et al., 2001; Mohan et al., 2013).

91 One way to adjust for A is to fit a “superpopulation model”, which is a regression (or other type of)
92 model for Y that includes A as covariates (Elliott & Valliant, 2017). In effect, the superpopulation
93 model stratifies sites based on levels of A and predicts the average or expected value of Y for each
94 stratum. If A explains the correlation between R and Y , the two should be uncorrelated within each
95 stratum, and the predicted within-stratum means should be close to their true values (as there is no
96 sampling bias). Averaging the within-stratum means of Y with respect to the distribution of A across
97 all sites in the landscape yields an “adjusted” estimate of the landscape-wide mean of Y (i.e. the target
98 quantity). More simply, the average of the superpopulation predictions across all sites is the estimate
99 of the landscape-wide mean—or, in practice, the average of the predictions for non-sampled sites and
100 the observations for sampled sites (Elliott & Valliant, 2017)—and it is unbiased if A completely
101 explains the geographic bias.

102 Causal diagrams and what to adjust for

103 To identify the variables in A , insight can be gleaned from causal inference, where a similar challenge
104 arises. Recall that A is the set of variables that explains the correlation between R and Y . In causal
105 inference, where the goal is to estimate the causal effect of one variable on another, analysts must
106 identify and adjust for the set of variables that explain the non-causal portion of the correlation
107 between the two (Pearl et al., 2016). One way to identify these variables, which could also be used to
108 identify the variables in A , is to construct “causal diagrams”(Greenland et al., 1999; Thoemmes &
109 Mohan, 2015).

110 Causal diagrams—not to be confused with parametric structural equation models—will not be
111 familiar to some ecologists (but see Grace & Irvine, 2020), so we will introduce the relevant concepts
112 using the example in Fig 1 (a full description of the graph in Fig. 1 is provided in the Methods
113 section). An arrow from one variable to another indicates a direct causal effect; that is, it indicates that
114 the cause is part of the real-world function that determines the value of the effect (Pearl et al., 2016).
115 In Fig. 1, annual temperature (annual_temp) has a direct effect on Y , which is a species’ abundance. A
116 path consists of several variables linked by arrows regardless of the direction of those arrows. Fig. 1
117 depicts a path linking R (site inclusion) to Y : $R \leftarrow$ heather_grass \rightarrow heather_grass_quality $\rightarrow Y$. The
118 existence of a path between two variables implies that they are correlated, whether by association or a
119 causal link. Hence, Fig. 1 implies a correlation between R and Y and therefore a geographic bias.



120

121 Figure 1. A simplistic causal diagram depicting causes and effects of sample inclusion R and a
122 species’ abundance Y .

123 While a path between two variables implies that they are correlated (dependent), it does not imply that
124 the dependence is unbreakable. Rather, it might be possible to “block” paths between variables, which
125 is to say, to block the flow of association. Blocking a path is achieved by adjusting for certain
126 variables (i.e. A), and a set of rules—the rules of d-separation—tell us which ones (Cinelli et al.,

127 2022; Pearl et al., 2016). It is not necessary to spell out the rules of d-separation here, partly because
128 they are described elsewhere (see the references above) and partly because they are built into software
129 packages such as the R package `dagitty` (Textor et al., 2016). The important point is that the rules
130 can be used to determine the sets of variables A that, when statistically adjusted for, will render R and
131 Y independent (if such a set exists; Thoemmes & Mohan, 2015).

132 Identifying the correct set of variables in A is contingent on the causal diagram being an accurate
133 reflection of reality, and there are broadly two ways to achieve this (which are not mutually
134 exclusive). One is to consult domain experts, who understand the relevant system, when constructing
135 the diagram (Grace & Irvine, 2020). The other is to develop the diagram iteratively and test the
136 implied conditional independencies (according to the rules of d-separation) of each iteration (Pearl et
137 al., 2016). Where data are missing on Y for most sites (i.e. where $R = 0$), it will not be possible to test
138 many of a diagram's implied conditional independencies, in which case consulting domain experts is
139 even more important than usual. Of course, experts cannot be expected to have a perfect
140 understanding of most systems, and we come back to this point in several places below.

141 Structure of this paper

142 In this paper, we demonstrate how one might combine expert knowledge with causal diagrams and
143 superpopulation models to mitigate geographic biases in biodiversity monitoring data. We assume that
144 a time trend in the mean of Y across all sites in the relevant landscape is the target quantity, and it is
145 estimated by fitting separate superpopulation models for each of several time-periods. Starting with
146 some simple simulations, we test the abilities of superpopulation models including A as covariates to
147 recover the landscape-wide mean of Y from perfect and imperfect (i.e. including measurement error)
148 data on A . Next, we apply our method to empirical data from the UK Butterfly Monitoring Scheme
149 (UKBMS). The scheme estimates time trends in the mean abundances butterflies in the UK using a
150 model that does not adjust for geographic bias, and we compare these trends to those produced using
151 superpopulation models. In the final section, we discuss the pros and cons of our approach and how it
152 could be improved in future.

153 Methods

154 Demonstrating superpopulation modelling via simulation

155 Although superpopulation models can mitigate geographic biases in theory, reality is more complex.
156 It is likely, for example, that data will not be available on some variable(s) in A at the exact location
157 of a site. Instead, the available data might represent an average across some larger area in which the
158 site is situated. Alternatively, data might be available on A at the precise location of the site but not
159 for the relevant time-period. In these cases, we would expect the data to correlate with the true
160 variable(s) in A , but not to be a perfect proxy. To test the implications of including imperfect data on
161 A in a superpopulation model, as well as some other methodological choices, we conducted a simple
162 simulation study. (See supplementary material three for an additional simulation that includes two
163 time-periods and the estimation of a trend.)

164 Our simulation was based on the causal diagram in Fig. 1. The diagram depicts the causes of some
165 species' abundance Y and sample inclusion R across 250,000 sites in a simple virtual landscape. It
166 implies that the quantity of good quality heather grassland (`heather_grass_quality`) and annual
167 temperature at each site directly affect Y . The quantity of good quality heather grassland at a site is
168 determined by the total quantity of heather grassland (of any quality; `heather_grass`) and the
169 proportion that is well managed (`management`). R is greater at sites with more heather grassland,
170 because recorders know that it is a favourable habitat for the focal species, which they hope to see.
171 The presence of major roads near a site makes it more accessible so also affect R . According to the
172 rules of d-separation, the causal diagram in Fig. 1 implies that R is independent of Y given heather
173 grassland (i.e. there is only variable in A).

174 To simulate the data, we used the `simulateSEM` function in the R package `dagitty` (Textor et al.,
175 2016). With the exception of R , all variables are standard normal (i.e. mean of 0 and unit variance).
176 Other than causes of R (see below), each variable explains ~25% of the variation in its direct

177 descendent. R is a binary variable (0 if the data is missing and 1 otherwise), so we could not simulate
 178 it in the same way as the others. Instead, following Thoemmes & Rose (2014), we simulated a latent
 179 standard normal variable and discretized it by assigning all values above the first percentile the value
 180 1 and the remainder the value 0 (i.e. we assume data on Y are missing for 99% of sites). A
 181 consequence of this strategy is that each cause of R explains less than 25% of its variance, because
 182 discretizing the latent normal variable attenuates the correlation between the two. Strictly speaking,
 183 the latent normal variable should be included in the causal diagram in Fig 1, but it does not affect its
 184 implied conditional independencies, so we omit it for simplicity.

185 Having simulated the data, we estimated the mean of Y across all sites using the data at sites where
 186 $R = 1$ (i.e. the sample). We used six estimators, which are listed in Table 1. The first is the sample
 187 mean, and the remainder are superpopulation models—in this case, linear regressions—including
 188 different covariates. The first superpopulation model includes A (heather_grass) as a covariate. The
 189 second includes A and an additional cause of Y (annual_temp), which should increase precision
 190 (Cinelli et al., 2022). The remainder of the superpopulation models include correlates of A (with
 191 different strengths of correlation) to reflect the fact that the available data are unlikely to be error-free.

192 Table 1. Six analytical approaches to estimating the mean of Y (species abundance) across all sites in
 193 the simulated landscape. The adjustment set is the set of covariates included in the superpopulation
 194 model.

Estimator	Adjustment set	Details
1	NA	Sample mean
2	heather_grass	Superpopulation model
3	heather_grass, annual_temp	Superpopulation model
4	correlate of heather_grass ($\rho = 0.5$)	Superpopulation model
5	correlate of heather_grass ($\rho = 0.7$)	Superpopulation model
6	correlate of heather_grass ($\rho = 0.9$)	Superpopulation model

195 We evaluated the superpopulation models' performances in terms of their estimation error. The
 196 procedure involved simulating 100 datasets based the causal diagram in Fig. 1 and calculating the
 197 difference between the true, landscape-wide mean of Y and the superpopulation model estimate for
 198 each one. The 100 datasets differed slightly from one another, as there is a random component to the
 199 `simulateSEM` function. Averaging the estimation error across the 100 estimates of mean Y for each
 200 method gives an estimate of its estimator bias.
 201

202 UKBMS case study

203 *The scheme*

204 To demonstrate how causal diagrams and superpopulation models might be used to correct biases in a
 205 real dataset, we applied them to data from the UK Butterfly Monitoring Scheme (UKBMS). The
 206 UKBMS has been running since 1976 (Pollard & Yates, 1996). Data are collected by volunteers, who
 207 walk transects at a network of sites in the UK and count the butterflies that they see within an
 208 imaginary 5-metre box when weather permits (Pollard, 1977). The volunteers are free to decide where
 209 to establish transects and generally do so in good quality semi-natural habitat, where butterflies are
 210 most abundant (Brereton, Roy, et al., 2011). Recorders are asked to walk UKBMS transects at least
 211 once in each of the 26 weeks from April 1st to September 29th to cover the main flight periods of UK
 212 butterflies. On average, however, ~20 weeks are sampled due to poor weather conditions, recorder
 213 availability etc.

214 In 2009, the Wider Countryside Butterfly Survey (WCBS) was established and incorporated in the
 215 UKBMS. The primary motivation for the WCBS was to increase coverage of habitats that were
 216 poorly represented by the UKBMS (Brereton, Cruickshanks, et al., 2011). Hence, rather than being
 217 chosen by the volunteers, WCBS sites are located within randomly selected 1 km grid squares in the
 218 UK. Importantly, however, volunteers are not willing/able to visit all WCBS squares: they prefer to
 219 visit and to re-visit sites that are accessible or where they are likely to see species that interest them.

220 Consequently, WCBS squares are not truly random. WCBS transects also differ from traditional
221 UKBMS transects in that volunteers are asked to walk them a minimum of twice in July–August
222 (although some do more) at least ten days apart.

223 *Existing analytical method*

224 The current method used to analyse the UKBMS (including WCBS) data has three steps (Dennis et
225 al., 2016). The first is to fit a GAM to estimate normalised seasonal abundance curves for each
226 species and year (these do not vary geographically; Dennis et al., 2013). The second is to estimate
227 annual site indices of relative abundance using the fitted GAMs. The total observed counts are scaled
228 by the proportion of the seasonal abundance curve that was surveyed to provide an estimate of the
229 expected total had the entire season been surveyed. Finally, a Poisson GLM with site and year effects
230 is fitted to the annual site indices, and the estimated year effects (log₁₀ scale) are used as annual
231 indices of relative abundance. In the final GLM, each site/year index is weighted in proportion to the
232 fraction of the 26-week flight curve that was sampled (rather than imputed). Consequently, the WCBS
233 sites, which are sampled as little as twice per year, should be heavily downweighted.

234 *New analytical method*

235 We modified the current UKBMS framework by replacing the final GLM with superpopulation
236 models that adjust for *A*. Steps one and two, which produce seasonally adjusted annual site indices of
237 relative abundance, remain the same, and the superpopulation models are fitted to these indices.
238 Taxon and dataset experts provided the information needed to construct causal diagrams and identify
239 the variables in *A* (more details below).

240 For demonstrative purposes, we focused on two species of butterfly, whose geographic distributions
241 and ecologies are very different: the meadow brown (*Maniola jurina*) and the small pearl-bordered
242 fritillary (*Boloria selene*). The causes and effects of *Y*, i.e. relative abundance, differ between the two
243 species. As we are only working with one dataset, collected by the UKBMS (including WCBS sites),
244 the causes and effects of *R* (site inclusion) do not differ between species.

245 We asked two taxon and UKBMS experts to provide information on the causes and effects of *R* and *Y*
246 via the forms in supplementary materials one and two. Both experts (redacted) are authors on this
247 paper. To constrain their answers, we asked the experts to select causes and effects from a list of 49
248 land cover (i.e. habitat), bioclimatic, geological and other variables on which data are available (see
249 supplementary material three for details of the data). Of course, this strategy risks omitting important
250 variables—a point we come back to in the Discussion. The experts provided their feedback
251 independently.

252 Rather than the transect-level, we asked the experts to think about causes and effects of *R* and *Y* at the
253 coarser resolution of 1 × 1 km, which was necessary for two reasons. First, it is the finest resolution
254 at which data on some variables in *A* are available. The second reason is more complex. Recall that
255 the superpopulation model predictions of *Y* must be averaged across all sites in the landscape. It
256 would be conceptually challenging the split the landscape into imaginary transects, which are the
257 “true” sites, but splitting it into 1 × 1 km grid squares is simple. In the few cases where multiple
258 UKBMS transects fell within the same grid square, we averaged the site indices across those transects
259 in a given year to obtain a grid square-level data point.

260 The information provided by the experts enabled us to construct causal diagrams and, using the rules
261 of d-separation, to identify the variables in *A* (according to each expert). In total, we produced four
262 causal diagrams: one for each species and expert. For any one diagram, there may be multiple sets of
263 variables *A* that d-separate *R* and *Y*. We selected the “minimal” (i.e. smallest) set that only included
264 variables on which data are available (listed in Table 2).

265 Both experts indicated that some land cover (habitat) types have negative effects on *R* and *Y*, but we
266 did not include these in the causal diagrams. Our logic is that these land cover types are only
267 detrimental to the species in the sense that they are not the right habitat (indeed, this is how it was
268 phrased by expert two [redacted]). The land cover data are expressed as the proportion of each 1 km
269 grid square covered by each land cover type (Morton et al., 2022), which means that a large

270 proportion of one necessarily means a small proportion of another (i.e. the land cover types are not
271 independent). That is, a high proportion of favourable habitat would imply a low proportion of
272 unfavourable habitat, and vice versa, so it is not necessary to include unfavourable land cover types in
273 the causal diagrams.

274 Having identified two sets of variables A (one from each expert) for each species, the next step was to
275 fit superpopulation models including these variables as covariates. We fitted one model per species,
276 expert (i.e. A) and year. As the data are (normalised) counts, and for consistency with the current
277 UKBMS method, we used Poisson GLMs with a log link function. We used each model to predict Y
278 for non-sampled sites, combined the predictions with the observations at sampled sites and calculated
279 the mean for the relevant species and year. Each year's mean for a given expert and species gives an
280 annual index of abundance. To measure change over time, we used the slope from a regression of the
281 estimated abundance indices on year.

282 For simplicity, we treated the model covariates (i.e. A) as constant over time. This was a natural
283 choice for elevation and calcareous bedrock, which are genuinely time-invariant (at least since 1976).
284 Our approach might seem less appropriate for the three land cover classes that featured in at least one
285 of the models (Table 2), but inspection of the available land cover maps shows that there has been
286 little change in the recent past (85% of 25×25 m grid squares in Great Britain were the same land
287 cover class in 2015 as in 1990; supplementary material three). Full details of the auxiliary data are
288 provided in supplementary material three.

289 *Comparing trends from the current and new methods*

290 Some rescaling was needed to ensure comparability between the outputs of the current UKBMS
291 analysis and the superpopulation models. The year effects from the current UKBMS model represent
292 the average change in the focal species' relative abundance across sites in the relevant year relative to
293 a reference year (1976) on the log₁₀ scale. The estimates from the superpopulation models, on the
294 other hand, reflect the average absolute count across sites in each year on the natural log scale. To
295 match the two, we converted the natural log estimates from the superpopulation models to the log₁₀
296 scale then subtracted the estimate from the reference year from each year's index.

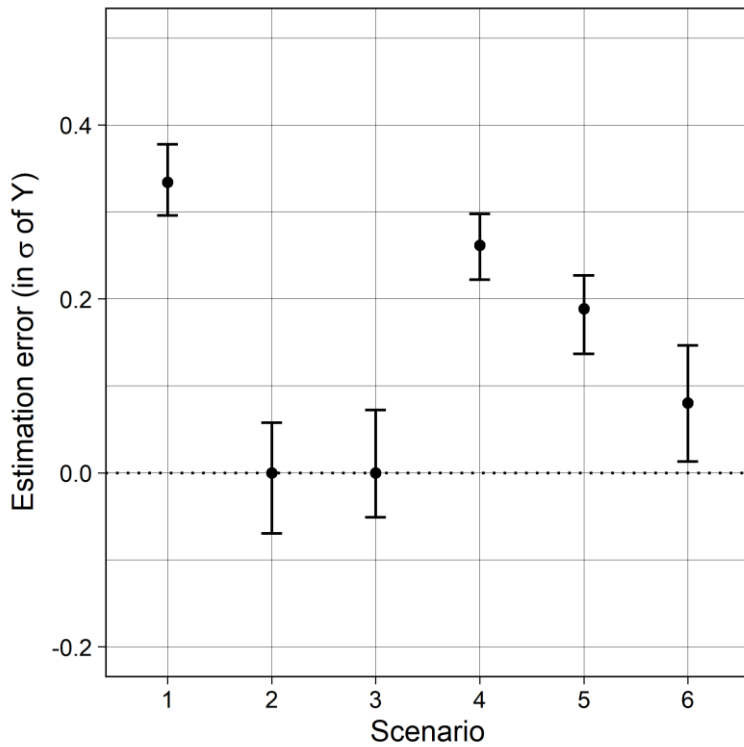
297 *Variance estimation*

298 For both the superpopulation models and the simpler GLM used by the UKBMS, we estimated the
299 sampling variance in each year's index of relative abundance by bootstrapping across sites. We
300 created 100 bootstrap samples by resampling the site indices with replacement within each year and
301 fitted a model to each sample. We present percentile (95%) confidence intervals for each year from
302 the bootstrap distribution of the annual indices (year effects from the simple GLM and estimates of
303 mean Y from the superpopulation models). Note that this procedure does not propagate uncertainty in
304 the estimation of the site indices from the raw count data.

305 **Results**

306 **Simulations**

307 Fig. 2 shows the estimation error—the difference between the true, landscape-wide mean of Y and the
308 estimate thereof—of each analytical method listed in Table 1. The sample mean is highly biased
309 (mean error, i.e. bias, of ~ 0.35 in units of standard deviations of Y). Using a superpopulation in which
310 A is the sole covariate (A comprises just one variable in this example), the estimate of mean Y across
311 all sites is unbiased (scenario 2; mean error ~ 0). The estimate is also unbiased if A plus an additional
312 cause of Y are included as covariates (scenario 3), and it has similar variance to the estimate from
313 scenario 2. Including a correlate of A as the covariate reduces the bias relative to the sample mean but
314 does not eliminate it (scenarios 4-6; see Table 1). The stronger the correlation between A and its
315 proxy, the closer the bias to that from scenario 2, where A itself was included as a covariate.



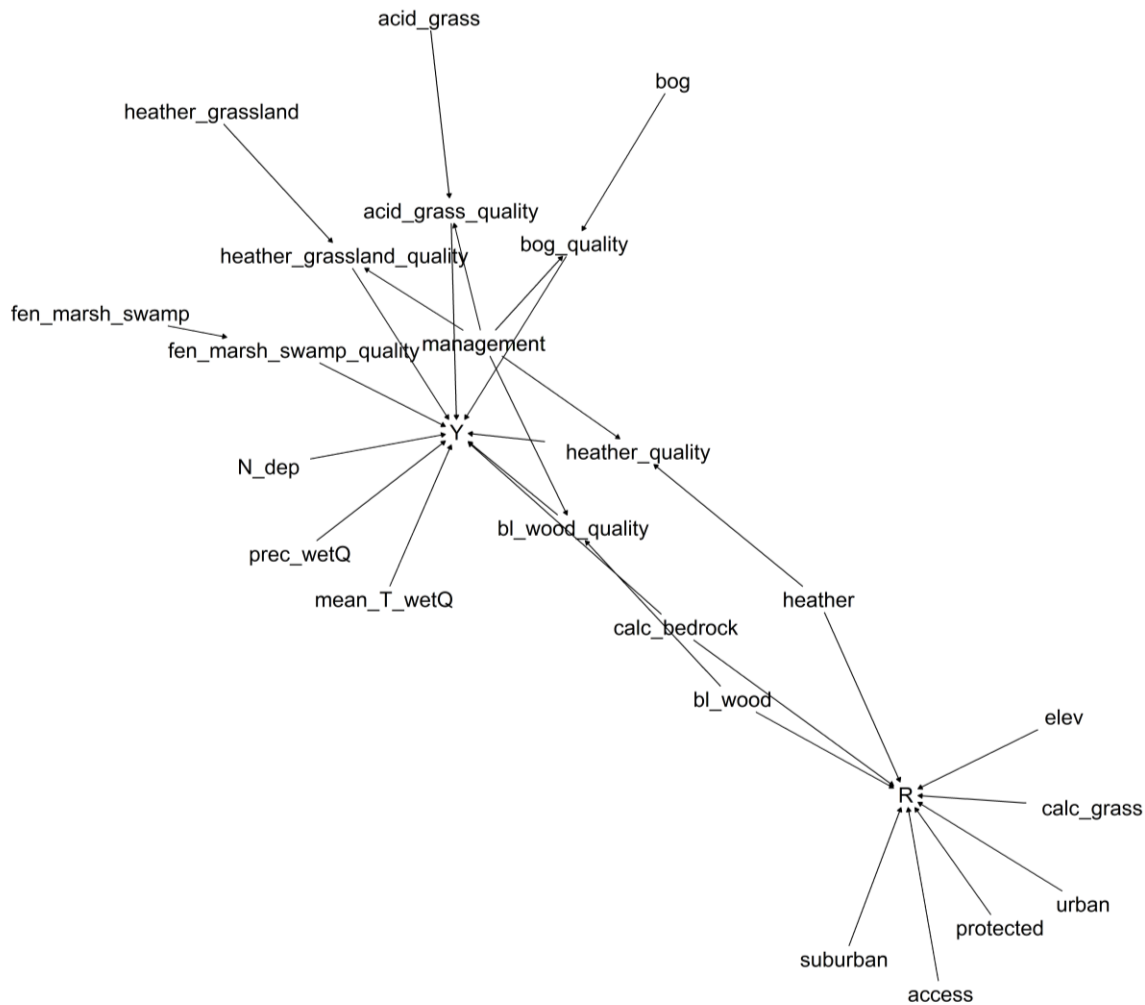
316

317 Figure 2. Estimation error (difference between the true, landscape-wide mean of Y and the estimate
 318 thereof) under each of the six scenarios in Table 1. The dots represent the mean error across 100
 319 simulated datasets, and the error bars depict the 2.5th and 97.5th percentiles.

320 UKBMS case study

321 *Identifying the variables in A*

322 The experts provided different perspectives on the causes and effects of R (UKBMS site inclusion
 323 including WCBS sites) and Y (the abundances of the meadow brown and small pearl-bordered
 324 fritillary). Their feedback is provided in full in supplementary materials one and two. Fig. 3 shows the
 325 causal diagram reflecting expert two's knowledge of the causes and effects of R and the Y in the case
 326 of the small pearl-bordered fritillary.



327

328 Figure 3. Causal diagram depicting causes and effects of sample inclusion in the UK Butterfly
 329 Monitoring Scheme (R) and the abundance of the small pearl-bordered fritillary (Y). The diagram
 330 reflects the knowledge of a taxon and UKBMS expert (expert two). N_dep is nitrogen deposition,
 331 $mean_T_wet_Q$ is the mean temperature of the wettest quarter of the year, $elev$ is elevation, bl_wood
 332 is broadleaved woodland, $protected$ is the proportion of each grid square that is some form
 333 of protected area and variables suffixed by “_quality” denote the proportion of each grid square that is
 334 both the relevant land cover type (indicated by the rest of its name) and managed appropriately for the
 335 species. The other variables are self-explanatory.

336 A major difference between the two experts’ feedback is that expert two indicated positive effects of
 337 several land cover classes (habitats) on Y conditional on the way that they are managed. That is, the
 338 habitats support high abundances of the species, but only if they are managed appropriately. We
 339 codified these conditional effects in the causal diagrams by introducing new, *unobserved* variables.
 340 One is “management”, which indicates whether the habitat is managed appropriately for the species,
 341 and the others are variables that denote the proportion of each site that is both the relevant land cover
 342 class and managed appropriately (i.e. “good” quality). For example, the expert indicated that heather
 343 grassland has a positive effect on the abundance of the small pearl-bordered fritillary conditional on
 344 appropriate management. In the causal diagram, the effect of heather grassland on Y is heather
 345 grassland \rightarrow quality heather grassland $\rightarrow Y$, where management also has a direct effect on quality
 346 heather grassland (Fig. 3). Note that we did not specifically ask the experts to provide information on
 347 conditional effects, which is why expert one did not.

348 From each of the four causal diagrams—one for each species and expert—we identified the minimal
 349 adjustment set A (the variables on which R and Y are hopefully conditionally independent) using the
 350 rules of d-separation (Table 2). For each diagram, it is possible to d-separate R and Y without having
 351 to adjust for any of the unobserved variables introduced to depict the conditional effects of land cover
 352 classes on Y (i.e. management or “quality” habitats). For both species, the variables in A differ
 353 between experts, which meant that two superpopulation models had to be fitted for each one.

354 Table 2. Minimal adjustment sets derived from the experts’ knowledge for each species. The
 355 adjustment set need not include all predictors of the species’ abundances.

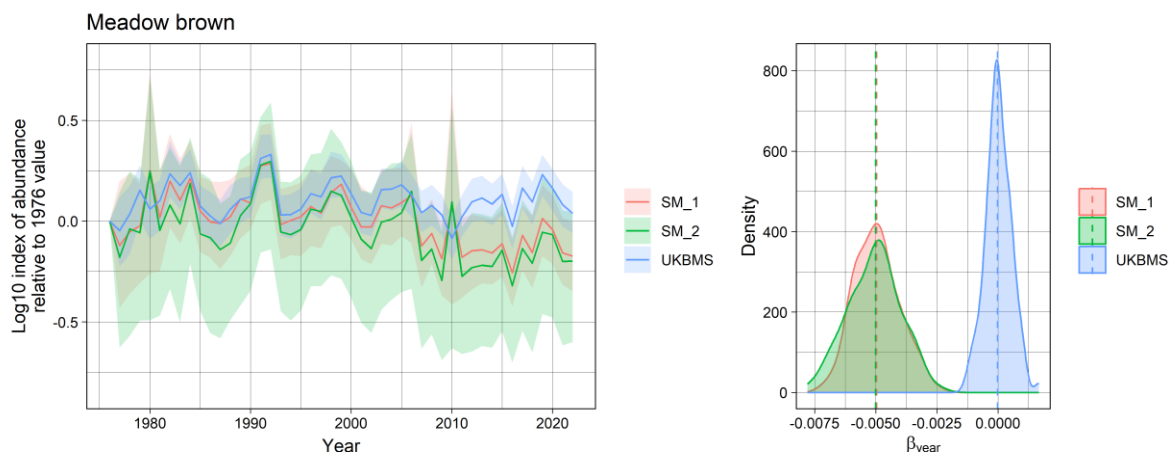
Expert	Species	Minimal adjustment set
1	meadow brown	Calcareous grassland
1	small pearl-bordered fritillary	Broadleaved woodland
2	meadow brown	Calcareous grassland and elevation
2	small pearl-bordered fritillary	Broadleaved woodland, calcareous bedrock and heather

356

357 *Current versus new model outputs*

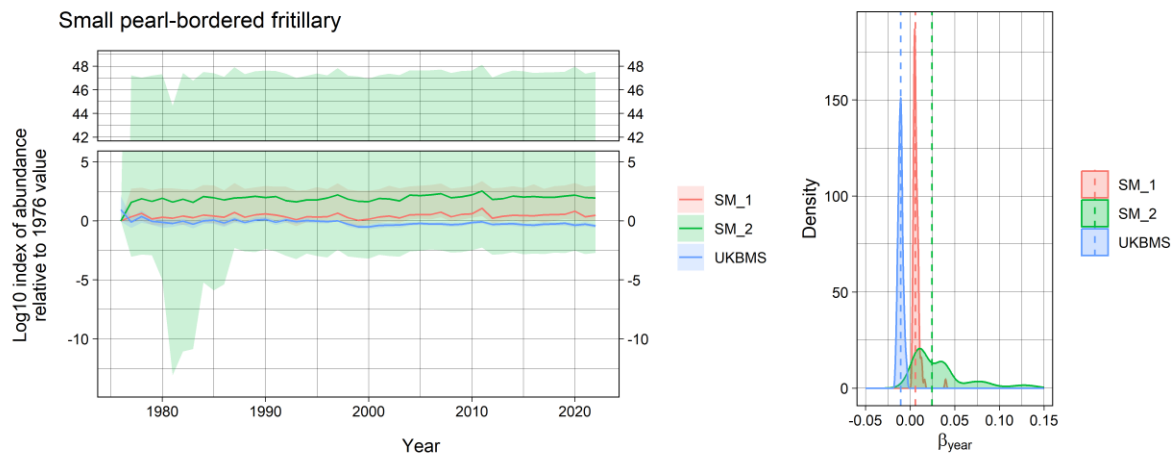
358 We compared the superpopulation models’ outputs to those of the current UKBMS GLM in terms of
 359 the annual estimates and the long-term trend, which is the coefficient from a regression of annual
 360 index on year (Figs 4 and 5). For both species, the outputs of the superpopulation models are
 361 markedly different to those of the UKBMS method. For the meadow brown, the two superpopulation
 362 models (one for each expert) agree that there has been a decline in abundance, whereas the UKBMS
 363 method suggests that abundance has been stable over time. For the small pearl-bordered fritillary, the
 364 current UKBMS method suggests a decline in abundance over time, whereas the two superpopulation
 365 models indicate a slight increase.

366 The estimates from one of the two superpopulation models (SM_2) for the small pearl-bordered
 367 fritillary appear to be very uncertain (Fig. 5). This uncertainty is a result of the fact that we rescaled
 368 the estimates in each year by subtracting the estimate from the reference year, 1976. In 1976, there
 369 were only eight grid squares with data on the small pearl-bordered fritillary, so the variance in this
 370 year is truly large. Subtracting these highly variable estimates from the estimates for subsequent years
 371 makes them appear similarly uncertain. The variance of the estimates on the absolute scale is far
 372 smaller (supplementary material three; Fig. S4), but we chose to present the estimates relative to 1976
 373 for consistency with the UKBMS approach.



374

375 Figure 4. Time trends in the annual indices of abundance for the meadow brown produced by the
 376 current UKBMS GLM (UKBMS), the superpopulation model based on expert one’s knowledge
 377 (SM_1) and the superpopulation model based on expert two’s knowledge (SM_2). The left panel
 378 shows the time-series, and the right panel shows the estimated trends, which are the coefficients from
 379 a regression of index value on year. The uncertainty in both panels was derived by bootstrapping
 380 across sites. The vertical lines in the right-hand panel represent the median slopes.



381

382 Figure 5. Time trends in the annual indices of abundance for the small pearl-bordered fritillary
 383 produced by the current UKBMS GLM (UKBMS), the superpopulation model based on expert one’s
 384 knowledge (SM_1) and the superpopulation model based on expert two’s knowledge (SM_2). The left
 385 panel shows the time-series, and the right panel shows the estimated trends, which are the coefficients
 386 from a regression of index value on year. The uncertainty in both panels was derived by bootstrapping
 387 across sites. The break in the Y axis on the left panel was introduced to show the full uncertainty in
 388 SM_2 whilst also enabling comparison of the three models. The vertical lines in the right-hand panel
 389 represent the median slopes.

390 Discussion

391 We have demonstrated how causal diagrams and superpopulation models might be used to estimate
 392 the landscape-wide mean of some variable of interest Y (e.g. a species’ abundance or occupancy) from
 393 geographically biased monitoring data. Our simulations show that, if the variables that explain the
 394 geographic bias (i.e. A) are measured (without error), known and included in the superpopulation
 395 model, then that model is unbiased (Fig. 2). Clearly, these ideal conditions are unlikely, a point we
 396 come back to below. By fitting models for multiple time-periods, we estimated time trends in the
 397 mean relative abundances of two species of butterfly using data from the UKBMS. The trends are
 398 different to those estimated using the UKBMS’s existing method, which does not adjust for
 399 geographic bias (Figs 4 and 5).

400 The experts (redacted) were not confident about whether the superpopulation models more accurately
 401 captured the UK-wide trends for the two species than the existing UKBMS method. The UKBMS
 402 conducts a regional breakdown of each species’ trend in the UK (<https://ukbms.org/official-statistics>).
 403 For the small pearl-bordered fritillary, it reports a strong decline in England but a strong increase in
 404 Scotland. Scotland is relatively underrepresented in the UKBMS data, so a good adjustment should
 405 place more weight on the trend in this region. The superpopulation models clearly give more weight
 406 to the Scottish trend than the UKBMS method—one even suggests that the species is increasing at the
 407 UK level. Nevertheless, two experts (redacted) are sceptical that the small pearl-bordered fritillary is
 408 not declining at the UK level (based on e.g. an observed decline in its distribution). All experts felt
 409 that the superpopulation model estimates for the meadow brown are plausible (although none were
 410 highly confident). Their reasoning is that the species is probably doing better or at least as well at
 411 sampled than at non-sampled locations.

412 From a theoretical perspective, there are some limited cases in which the UKBMS model and others
 413 like it could capture the true trend in the landscape-wide mean of a species’ abundance Y despite not
 414 adjusting for geographic biases. If, for example, the bias is time-invariant in terms of both sign and
 415 magnitude, then no adjustment is needed. This scenario is highly unlikely, however. The distribution
 416 of sample inclusion R changes over time, because different sites are sampled in different years. So too
 417 does the distribution of Y unless the focal species’ relative abundance remains the same at each site in
 418 each time-period, which is clearly implausible. As the distributions of R and Y change over time, it is

419 highly unlikely that the correlation between the two, i.e. the bias, would remain constant. We also
420 note that the existing UKBMS model could produce unbiased estimates of alternative estimands using
421 the UKBMS data. For example, one might be interested in the trend in the mean abundance across
422 occupied sites (i.e. where $Y \geq 1$), in which case sampling all populations would permit unbiased
423 inference. Of course, this would require knowing the locations of all populations, and range shifts
424 would complicate matters.

425 The major benefit of using superpopulation models rather than conventional methods (e.g. a GLM
426 with site and year effects) is that they relax the highly untenable assumption that there is no
427 geographic bias; instead, they work on the weaker and slightly more tenable assumption that there is
428 no geographic bias conditional on the adjustment set A . Of course, satisfying this weaker assumption
429 means being confident that i) experts were able to identify all relevant causes and effects of sample
430 inclusion R and the variable of interest Y and that ii) data on those variables are available, which are
431 doubtful. The hope is that the models have reduced the geographic bias, i.e. the correlation between R
432 and Y , to the point where the adjusted estimates of the landscape-wide mean of Y are appreciably less
433 biased than the naïve ones.

434 Whether a superpopulation model estimate of the landscape-wide mean of Y is less biased than a
435 naïve (i.e. unadjusted) one depends on several factors, some of which were captured by our
436 simulations. Where data on all A are available, but they were measured with error, the adjustment
437 should reduce but not eliminate bias (Fig. 2). Matters are more complex where some variables in A
438 are omitted from the model and/or where some variables that are not in A are erroneously included
439 (both of which may result from mis-specifying the causal diagram). Erroneously included/omitted
440 variables that are more strongly related to R and Y have greater potential to cause a bias, because they
441 can induce a larger correlation between the two (Collins et al., 2001; Thoemmes & Rose, 2014).
442 However, the sign of the relationships is also relevant: if one erroneously omitted/included variable
443 induces a negative geographic bias and another induces a positive one, then the two might cancel each
444 other out and cause no bias (Thoemmes & Rose, 2014). Our simulations do not capture the effects of
445 mis-specifying the causal diagrams, because any choice of a select few scenarios of strengths and
446 directions of effects and causal diagram structures would have been necessarily arbitrary, but analysts
447 should recognise that the success of the superpopulation modelling approach is contingent on
448 accuracy of the causal diagram.

449 Indeed, constructing accurate causal diagrams is the major practical limitation of our approach.
450 Analysts might have hundreds or even thousands of species in mind, and experts might not have the
451 knowledge to construct causal diagrams for each one. Even if they did, it would be a time-consuming
452 exercise (although less costly than full probability sampling, which would be needed to satisfy the
453 assumptions of conventional methods). One option for reducing this burden might be to identify and
454 adjust for a set of variables that have large causal effects on R and Y for many species: for example,
455 woodland in the case of woodland birds or grassland in the case of grassland butterflies. Such a
456 strategy is likely to miss important variables in A for some species, however, so it should only be used
457 as a last resort. Another option might be to crowd-source information on the causes and effects of R
458 and Y from taxon and dataset experts. This strategy could work well for well-studied taxa such as
459 butterflies, birds and mammals, but it will be more challenging for species whose autecology is less
460 understood. It might also yield feedback from more than two experts per species (as here), which
461 would be desirable.

462 Another question is how comprehensive the causal diagrams need to be. It might be sufficient to
463 include direct causes of R and Y and the causal links among those variables as we did here. However,
464 there is a risk that including additional variables in the diagram, which do not do not directly affect R
465 or Y , might reveal paths between them. For example, imagine the diagram broadleaved woodland \rightarrow
466 R ; heather grassland $\rightarrow Y$. According to this diagram, R and Y are independent, and there is no
467 geographic bias. Now suppose that we construct a more comprehensive diagram that includes annual
468 temperature as a cause of both broadleaved woodland (say, density) and heather grassland. Extending
469 the diagram in this way reveals a path (i.e. a correlation) between Y and R that would have been
470 missed by including only direct causes of the two, and it would change the adjustment set A . Omitting

471 variables whose effects on R and Y are indirect is less problematic than omitting proximal ones,
472 because indirect causes necessarily explain less of the variation in the two (Grace & Irvine, 2020).
473 Nevertheless, analysts should acknowledge that a more comprehensive causal diagram is more likely
474 to yield the true set of variables in A , and this must be balanced against the increased effort required to
475 construct it (also noting that a more comprehensive diagram might result in a more complex
476 superpopulation model, which could increase its variance).

477 Having identified the variables in A , superpopulation models are just one of several methods that
478 could be used to adjust for them (Boyd, Stewart, et al., 2023). Alternatives include inverse probability
479 weighting (Fink et al., 2023; Johnston et al., 2020), poststratification (Van Swaay et al., 2002) and
480 others. Each of these methods can be recast mathematically as a means to eliminating the correlation
481 between R and Y (Meng, 2022), but, crucially, they are only guaranteed to achieve this goal where the
482 variables in A have been identified correctly (Bailey, 2022).

483 Alternative methods exist that do not require complete knowledge of or data on the variables in A , but
484 their assumptions are challenging if not impossible to evaluate. It is becoming common to model
485 geographic bias using spatial random fields, for example, which assign a value of sampling intensity
486 to every location in the landscape based on spatial autocorrelation in the sampling locations (e.g.
487 Simmonds et al., 2020). Use of this approach essentially substitutes the assumption that R and Y are
488 independent given A for the assumption that they are independent given the spatial field and other
489 covariates in the model (Diggle et al., 2010). Whether this assumption holds depends on whether the
490 true variables in A exhibit a consistent pattern of spatial autocorrelation across the landscape, which is
491 not possible to verify without first identifying those variables.

492 It might seem like using causal diagrams and superpopulation models to estimate time trends in
493 species' abundances requires stronger assumptions than traditional approaches (e.g. the GLM used by
494 the UKBMS). Importantly, however, these non-parametric assumptions are codified explicitly in the
495 form of causal diagrams (although parametric assumptions such as linear effects of covariates are
496 not), whereas the assumptions of simpler approaches are often overlooked. When the goal is to draw
497 landscape-scale inferences, using a model that does not adjust for geographic bias means assuming
498 that it does not exist. Not only is this assumption untenable (we know that the UKBMS' and other
499 biodiversity data are geographically biased), but it has the potential to introduce substantial estimation
500 error if violated. As they make different assumptions, superpopulation models can be fitted in addition
501 to conventional models as part of a sensitivity analysis (as in this paper). They can also be used as the
502 primary method of analysis where there is clear risk of bias (Boyd et al., 2022).

503 **Acknowledgements:** RJB and OLP were supported by the NERC Exploring the Frontiers award
504 number NE/X010384/1 "Biodiversity indicators from nonprobability samples: Interdisciplinary
505 learning for science and society". OLP was also supported by the NERC award number
506 NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability. The UK
507 Butterfly Monitoring Scheme is organized and funded by Butterfly Conservation, the UK Centre for
508 Ecology & Hydrology, British Trust for Ornithology, and the Joint Nature Conservation Committee.
509 The UKBMS is indebted to all volunteers who contribute data to the scheme.

510 **Data availability:** The butterfly data must be requested from the UK Butterfly Monitoring Scheme at
511 <https://ukbms.org/request-data>. All code used to conduct our analyses are provided in the
512 supplementary materials. We will deposit the code on Zenodo should the article be accepted for
513 publication.

514 References

- 515 Bailey, M. A. (2023). A New Paradigm for Polling. *Harvard Data Science Review*, 5(3).
516 <https://doi.org/10.1162/99608f92.9898eede>
- 517 Bailey, Michael, A. (2022). Comments on "Statistical inference with non-probability survey samples
518 ." *Survey Methodology*, 48(12), 331–338.

- 519 Bowler, D. E., Bhandari, N., Repke, L., Beuthner, C., Callaghan, C. T., Eichenberg, D., Henle, K.,
520 Klenke, R., Richter, A., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Decision-making of
521 citizen scientists when recording species observations. *Scientific Reports*, *12*(1), 1–12.
522 <https://doi.org/10.1038/s41598-022-15218-2>
- 523 Boyd, R. J., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin,
524 G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J., & Pescott, O. L. (2022). ROBITT: A
525 tool for assessing the risk-of-bias in studies of temporal trends in ecology. *Methods in Ecology
526 and Evolution*, *13*(March), 1497–1507. <https://doi.org/10.1111/2041-210X.13857>
- 527 Boyd, R. J., Powney, G. D., & Pescott, O. L. (2023). We need to talk about nonprobability samples.
528 *Trends in Ecology & Evolution*, *38*(6), 521–531. <https://doi.org/10.1016/j.tree.2023.01.001>
- 529 Boyd, R. J., Stewart, G. B., & Pescott, O. L. (2023). Descriptive inference using large,
530 unrepresentative nonprobability samples: An introduction for ecologists. *Ecology*.
531 <https://doi.org/10.1002/ecy.4214>
- 532 Brereton, T., Cruickshanks, K. L., Risely, K., Noble, D. G., & Roy, D. B. (2011). Developing and
533 launching a wider countryside butterfly survey across the United Kingdom. *Journal of Insect
534 Conservation*, *15*(1), 279–290. <https://doi.org/10.1007/s10841-010-9345-8>
- 535 Brereton, T., Roy, D. B., Middlebrook, I., Botham, M., & Warren, M. (2011). The development of
536 butterfly indicators in the United Kingdom and assessments in 2010. *Journal of Insect
537 Conservation*, *15*(1), 139–151. <https://doi.org/10.1007/s10841-010-9333-z>
- 538 Cinelli, C., Forney, A., & Pearl, J. (2022). A Crash Course in Good and Bad Controls. *Sociological
539 Methods and Research*, *March*, 1–30. <https://doi.org/10.1177/00491241221099552>
- 540 Collins, L. M., Schafer, J., & Kam, C. (2001). A Comparison of Restrictive Strategies in Modern
541 Missing Data Procedures. *Psychological Methods*, *6*(June). <https://doi.org/10.1037/1082-989X.6.4.330>
- 543 Dennis, E. B., Freeman, S. N., Brereton, T., & Roy, D. B. (2013). Indexing butterfly abundance whilst
544 accounting for missing counts and variability in seasonal pattern. *Methods in Ecology and
545 Evolution*, *4*(7), 637–645. <https://doi.org/10.1111/2041-210X.12053>
- 546 Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T. M., & Roy, D. B. (2016). A generalized
547 abundance index for seasonal invertebrates. *Biometrics*, *72*(4), 1305–1314.
548 <https://doi.org/10.1111/biom.12506>
- 549 Diggle, P. J., Menezes, R., & Su, T.-L. (2010). Geostatistical inference under preferential sampling. In
550 *Appl. Statist* (Issue 2). <http://www.blackwellpublishing.com/rss>
- 551 Elliott, M. R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, *32*(2),
552 249–264. <https://doi.org/10.1214/16-STS598>
- 553 Fink, D., Johnston, A., Auer, M. T., Hochachka, W. M., Ligocki, S., Oldham, L., Robinson, O.,
554 Wood, C., Kelling, S., Rodewald, A. D., & Fink, D. (2023). A Double machine learning trend
555 model for citizen science data. *Methods in Ecology and Evolution*, *2023*(June), 1–14.
556 <https://doi.org/10.1111/2041-210X.14186>
- 557 Forister, M. L., Black, S. H., Elphick, C. S., Grames, E. M., Halsch, C. A., Schultz, C. B., & Wagner,
558 D. L. (2023). Missing the bigger picture: Why insect monitoring programs are limited in their
559 ability to document the effects of habitat loss. *Conservation Letters*, *September 2022*, 1–6.
560 <https://doi.org/10.1111/conl.12951>
- 561 Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C.,
562 & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four
563 recording schemes with different proficiency requirements. *Diversity and Distributions*, *22*(11),
564 1139–1149. <https://doi.org/10.1111/ddi.12477>

- 565 Grace, J. B., & Irvine, K. M. (2020). Scientist's guide to developing explanatory statistical models
566 using causal analysis principles. *Ecology*, *101*(4), 1–14. <https://doi.org/10.1002/ecy.2962>
- 567 Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. In
568 *Epidemiology* (Vol. 10, Issue 1, pp. 37–48). [https://doi.org/10.1097/00001648-199901000-](https://doi.org/10.1097/00001648-199901000-00008)
569 00008
- 570 Hughes, A., Orr, M., Ma, K., Costello, M., Waller, J., Provoost, P., Zhu, C., & Qiao, H. (2020).
571 Sampling biases shape our view of the natural world. *Ecography*, *44*, 1259–1269.
572 <https://doi.org/10.1111/ecog.05926>
- 573 Johnston, A., Moran, N., Musgrove, A., Fink, D., & Baillie, S. R. (2020). Estimating species
574 distributions from spatially biased citizen science data. *Ecological Modelling*, *422*(December
575 2019), 108927. <https://doi.org/10.1016/j.ecolmodel.2019.108927>
- 576 Lohr, S. (2022). *Sampling: Design and analysis* (3rd ed.). CRC Press.
- 577 Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big
578 data paradox, and the 2016 us presidential election. *Annals of Applied Statistics*, *12*(2), 685–726.
579 <https://doi.org/10.1214/18-AOAS1161SF>
- 580 Meng, X.-L. (2022). Comments on the Wu (2022) paper by Xiao-Li Meng 1 : Miniaturizing data
581 defect correlation : A versatile strategy for handling non-probability samples. *Survey*
582 *Methodology*, *48*(2), 1–22.
- 583 Mohan, K., Pearl, J., & Tian, J. (2013). Graphical models for inference with missing data. *Advances*
584 *in Neural Information Processing Systems, December*, 1277–1285.
- 585 Morton, R., Marston, C., O'Neil, A., & Rowland, C. (2022). *Land Cover Map 2018 (1km summary*
586 *rasters, GB and N. Ireland)*. NERC EDS Environmental Information Data Centre.
587 <https://doi.org/https://doi.org/10.5285/9b68ee52-8a95-41eb-8ef1-8d29e2570b00>
- 588 Moussy, C., Burfield, I. J., Stephenson, P. J., Newton, A. F. E., Butchart, S. H. M., Sutherland, W. J.,
589 Gregory, R. D., McRae, L., Bubb, P., Roesler, I., Ursino, C., Wu, Y., Retief, E. F., Udin, J. S.,
590 Urazaliyev, R., Sánchez-Clavijo, L. M., Lartey, E., & Donald, P. F. (2022). A quantitative global
591 review of species population monitoring. *Conservation Biology*, *36*(1), 1–14.
592 <https://doi.org/10.1111/cobi.13721>
- 593 Pearl, J., Glymour, M., & Jewell, N. (2016). *Causal inference in statistics: A primer*. Wiley.
- 594 Pescott, O. L., Walker, K. J., Harris, F., New, H., Cheffings, C. M., Newton, N., Jitlal, M., Redhead,
595 J., Smart, S. M., & Roy, D. B. (2019). The design, launch and assessment of a new volunteer-
596 based plant monitoring scheme for the United Kingdom. *PLoS ONE*, *14*(4), 1–30.
597 <https://doi.org/10.1371/journal.pone.0215891>
- 598 Pescott, O. L., Walker, K. J., Pocock, M. J. O., Jitlal, M., Outhwaite, C. L., Cheffings, C. M., Harris,
599 F., & Roy, D. B. (2015). Ecological monitoring with citizen science: The design and
600 implementation of schemes for recording plants in Britain and Ireland. *Biological Journal of the*
601 *Linnean Society*, *115*(3), 505–521. <https://doi.org/10.1111/bij.12581>
- 602 Pollard, E. (1977). A method for assessing changes in the abundance of butterflies. *Biological*
603 *Conservation*, *12*(2), 115–134. [https://doi.org/10.1016/0006-3207\(77\)90065-9](https://doi.org/10.1016/0006-3207(77)90065-9)
- 604 Pollard, E., & Yates, T. J. (1996). *Monitoring Butterflies for Ecology and Conservation*. Chapman &
605 Hall.
- 606 Powney, G. D., Carvell, C., Edwards, M., Morris, R. K. A., Roy, H. E., Woodcock, B. A., & Isaac, N.
607 J. B. (2019). Widespread losses of pollinating insects in Britain. *Nature Communications*,
608 *10*(2019), 1–6. <https://doi.org/10.1038/s41467-019-08974-9>

- 609 Robbins, C., Bystrak, D., & Geissler, P. (1986). *The Breeding Bird Survey: Its First Fifteen*
610 *Years, 1965-1979*. <https://pubs.usgs.gov/publication/5230189>
- 611 Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B., & Hara, R. B. O. (2020). Is more data
612 always better? A simulation study of benefits and limitations of integrated distribution models.
613 *Ecography*, *43*, 1413–1422. <https://doi.org/10.1111/ecog.05146>
- 614 Stroh, P. A., Walker, K., Humphrey, T. A., Pescott, O. L., & Burkmar, R. (2023). *Plant Atlas 2020:*
615 *Mapping Changes in the Distribution of the British and Irish Flora*. Princeton Univ. Press.
- 616 Textor, J., van der Zander, B., Gilthorpe, M. S., Liškiewicz, M., & Ellison, G. T. (2016). Robust
617 causal inference using directed acyclic graphs: The R package “dagitty.” *International Journal*
618 *of Epidemiology*, *45*(6), 1887–1894. <https://doi.org/10.1093/ije/dyw341>
- 619 Thoemmes, F., & Mohan, K. (2015). Graphical Representation of Missing Data Problems. *Structural*
620 *Equation Modeling*, *22*(4), 631–642. <https://doi.org/10.1080/10705511.2014.937378>
- 621 Thoemmes, F., & Rose, N. (2014). A Cautious Note on Auxiliary Variables That Can Increase Bias in
622 Missing Data Problems. *Multivariate Behavioral Research*, *49*(5), 443–459.
623 <https://doi.org/10.1080/00273171.2014.931799>
- 624 Van Swaay, C. A. M., Plate, C. L., & Van Strien, A. J. (2002). Monitoring butterflies in the
625 Netherlands: how to get unbiased indices. *Proceedings of the Section Experimental and Applied*
626 *Entomology of the Netherlands Entomological Society*, *13*, 21–27.
- 627 Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and
628 time. *Trends in Ecology and Evolution*, *16*(8), 446–453. [https://doi.org/10.1016/S0169-](https://doi.org/10.1016/S0169-5347(01)02205-4)
629 [5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)
- 630