



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Integrated Population Models: Achieving their Potential

**Citation for published version:**

Frost, F, McCrea, R, King, R, Gimenez, O & Zipkin, E 2022, 'Integrated Population Models: Achieving their Potential', *Journal of Statistical Theory and Practice*, vol. 17, 6. <https://doi.org/10.1007/s42519-022-00302-7>

**Digital Object Identifier (DOI):**

[10.1007/s42519-022-00302-7](https://doi.org/10.1007/s42519-022-00302-7)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Statistical Theory and Practice

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Integrated Population Models: Achieving their Potential

Fay Frost<sup>1\*</sup>, Rachel McCrea<sup>2</sup>, Ruth King<sup>3</sup>, Olivier Gimenez<sup>4</sup>  
and Elise Zipkin<sup>5</sup>

<sup>1</sup>\*School of Mathematics, Statistics and Actuarial Science,  
University of Kent.

<sup>2</sup>Department of Mathematics and Statistics, Lancaster University.

<sup>3</sup>School of Mathematics and Maxwell Institute for Mathematical  
Sciences, University of Edinburgh.

<sup>4</sup>CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier,  
France.

<sup>5</sup>Department of Integrative Biology, Ecology, Evolution, and  
Behavior Program, Michigan State University.

\*Corresponding author(s). E-mail(s): [F.Frost@kent.ac.uk](mailto:F.Frost@kent.ac.uk);  
Contributing authors: [R.McCrea@lancaster.ac.uk](mailto:R.McCrea@lancaster.ac.uk);  
[Ruth.King@ed.ac.uk](mailto:Ruth.King@ed.ac.uk); [olivier.gimenez@cefe.cnrs.fr](mailto:olivier.gimenez@cefe.cnrs.fr);  
[ezipkin@msu.edu](mailto:ezipkin@msu.edu);

## Abstract

Precise and accurate estimates of abundance and demographic rates are primary quantities of interest within wildlife conservation and management. Such quantities provide insight into population trends over time and the associated underlying ecological drivers of the systems. This information is fundamental in managing ecosystems, assessing species conservation status and developing and implementing effective conservation policy. Observational monitoring data are typically collected on wildlife populations using an array of different survey protocols, dependent on the primary questions of interest. For each of these survey designs, a range of advanced statistical techniques have been developed which are typically well understood. However, often multiple types of data may exist for the same population under study. Analysing each

31 data set separately implicitly discards the common information contained  
32 in the other data sets. An alternative approach that aims to optimise  
33 the shared information contained within multiple data sets is to use a  
34 “model-based data integration” approach, or more commonly referred to  
35 as an “integrated model”. This integrated modeling approach simultane-  
36 ously analyses all the available data within a single, and robust, statistical  
37 framework. This paper provides a statistical overview of ecological inte-  
38 grated models, with a focus on integrated population models (IPMs)  
39 which include abundance and demographic rates as quantities of interest.  
40 Four main challenges within this area are discussed, namely model spec-  
41 ification, computational aspects, model assessment and forecasting. This  
42 should encourage researchers to explore further and develop new practical  
43 tools to ensure that full utility can be made of IPMs for future studies.

44 **Keywords:** abundance, ecological insight, integrating data, multiple surveys

## 45 1 Introduction

46 A key goal in the study of wildlife populations is often to estimate abundance  
47 and important demographic rates (e.g. recruitment and survival) of species  
48 and how these variables change over space and time. Accurate and precise esti-  
49 mates of such quantities lay the foundation of determining abundance trends  
50 and the ecological dynamics of species and thus are necessary for effective con-  
51 servation planning and management in the face of ongoing global change [1].  
52 For example, inferences demonstrating changes in a population’s abundance,  
53 and the mechanisms behind such change, can aid in decisions on how to halt  
54 declines or manage the invasion of deleterious species [2–4]. As inferences on  
55 wildlife population parameters need to take into account a variety of processes,  
56 including imperfect detection of species, extreme heterogeneity in and among  
57 environments, and the movement and clustering of individuals, a variety of  
58 data collection and analysis frameworks have been developed over the last  
59 century to provide the relevant and necessary metrics on wildlife populations  
60 [5–7].

61 A range of data are collected on wild populations with the specific bio-  
62 logical questions, as well as logistical constraints, shaping the distinct types  
63 of data that are collected. Often, several types of data are collected on a sin-  
64 gle population or species within close proximity, as different researchers may  
65 be interested in multiple aspects of a particular study system. Historically,  
66 each data set would be analysed separately, with perhaps estimates from one  
67 analysis being used in another or biological interpretations compared, again  
68 depending on the questions of interest and data types available. However, sepa-  
69 rate analyses of available data sources discard valuable information that could  
70 improve the estimation of the biological quantities of interest, for example, by  
71 increasing the precision of parameter estimates [8], permitting the estimation  
72 of confounded parameters [9], and/or correcting for sampling biases in one or  
73 more data sources [10].

74 An alternative approach, which optimises the shared information contained  
75 within multiple data sets, is to use “model-based data integration”. Model-  
76 based data integration (or data integration, for short) is an umbrella term  
77 that refers to any modelling technique that simultaneously analyse all available  
78 data within a single, robust, statistical framework (e.g., data fusion, inte-  
79 grated data models). The development of data integration methods over the  
80 last three decades has grown almost exponentially [11–13], as these approaches  
81 can greatly reduce uncertainty of parameter estimates [14], make possible the  
82 estimation of parameters that are inestimable from a single data source (e.g.,  
83 emigration and mortality [15]), and expand the spatiotemporal scope of infer-  
84 ence [16]. Of note is that an individual data set may provide no additional  
85 information on the particular quantity of primary interest; however, by pro-  
86 viding direct information on other model parameters, this can, in turn, lead to

4 *Integrated Population Models: Achieving their Potential*

87 the ability to estimate a previously confounded parameter and/or an improved  
88 estimate of the quantity of interest [17].

89 In this paper, the ideas associated with integrated modelling are described  
90 and several associated outstanding statistical challenges discussed, with a spe-  
91 cific focus on integrated population models (IPMs). While data integration  
92 approaches aim to estimate a variety of population processes (e.g., species  
93 distributions, abundances), IPMs specifically focus on the simultaneous esti-  
94 mation of both population abundance and demographic rates within a single  
95 analytical framework. IPMs may integrate across many types of data, typ-  
96 ically including (but not limited to): 1) abundance, counts, and/or census  
97 data to inform temporally varying population sizes and/or trends, 2) produc-  
98 tivity data to inform (annual) reproduction rates (e.g., nest records), and 3)  
99 capture-recapture-recovery-type data used to inform (seasonal, annual, and/or  
100 stage/age) survival.

101 [1, 18] provide an overview of common types of available data included in  
102 IPMs and associated component models that may be applied. Such models may  
103 be improved via the inclusion of environmental covariate data [19, 20]. IPMs  
104 have also recently been developed combining data sets (possibly the same type  
105 of data) across different species [21–23], although this work is fairly new and  
106 still under development.

107 The first use of integrated approaches within ecology were applied to stock  
108 assessments within the fisheries industry [12]. However, IPMs gained promi-  
109 nence in the wider ecological community over two decades ago [7, 19, 24]  
110 when the approaches were applied to terrestrial species, usually birds. This  
111 early work combines state-space modelling of census data with the analysis  
112 of capture-recapture-recovery-type data, exploiting the shared demographic  
113 parameters between the two modelling techniques. There has been substantial

114 advancement over the last decade in particular expanding analyses to popula-  
115 tions of birds, mammals, and amphibians [25]. However, widespread adoption  
116 of IPMs by the ecological community is hindered by technical statistical and  
117 implementation challenges. The aim of this paper is to provide a statistical  
118 overview of IPMs, including the identification of many outstanding statistical  
119 challenges within this area, which in turn will encourage researchers to inves-  
120 tigate these issues and develop innovative and practical approaches to ensure  
121 that IPMs are able to reach their full potential within ecological studies.

## 122 **2 Integrated Data**

123 The fundamental concept of an integrated data approach is to estimate the  
124 ecological parameters of interest using all available information within a single  
125 and robust analysis. In particular, it is envisaged that there are multiple data  
126 sources that each provide information on the given ecological system/popula-  
127 tion of interest. For example, this may relate to multiple data sets collected on  
128 the same species but using different data collection survey techniques, or data  
129 sets relating to (potentially interacting) multiple species within the same geo-  
130 graphical location, or even data sets that are separated geographically and/or  
131 temporally. In all cases, it is assumed that the different data sets provide  
132 information, either directly or indirectly, on some mechanism within the given  
133 ecosystem of interest.

134 Mathematically, suppose there are  $n$  distinct data collection surveys, lead-  
135 ing to associated data sets  $\mathbf{x}_i$ , for  $i = 1, \dots, n$ . Combining these data sets  
136 leads to the integrated data  $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ . Given an associated model for  
137 these data, the associated global likelihood of the observed data is written as  
138  $f_G(\mathbf{x}; \boldsymbol{\theta})$ , here  $\boldsymbol{\theta}$  denotes the set of all model parameters across the different

139 data collection processes. The form of this likelihood is now considered in more  
140 detail, assuming different dependent structures for the observed data.

## 141 **2.1 Conditionally independent data**

142 Assuming that the different data sets are independent of each other, condi-  
143 tional on the associated model parameters, the global likelihood,  $f_G(\mathbf{x}; \boldsymbol{\theta})$  can  
144 simply be expressed as,

$$145 \quad f_G(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^n f_i(\mathbf{x}_i; \boldsymbol{\theta}),$$

146

147 where  $f_i(\mathbf{x}_i; \boldsymbol{\theta})$  denotes the likelihood of the observed data  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ . In  
148 general, this substantially simplifies the model specification, since each likeli-  
149 hood is constructed independently for each data set. In practice, it is assumed  
150 that there are some model parameters that are common across the different  
151 data sets, motivating an integrated data analysis approach. Conversely, if the  
152 model parameters are non-overlapping across the different data collection pro-  
153 cesses, so that  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n\}$ , where  $\boldsymbol{\theta}_i \cap \boldsymbol{\theta}_j = \emptyset$  for  $i \neq j$ , the likelihood  
154 reduces to,

$$155 \quad f(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta}_i).$$

156

157 In this special case, analysing the joint likelihood is equivalent to simply  
158 analysing each data set independently of each other, and there is no benefit  
159 gained in considering an integrated approach. Thus, for the remainder of the  
160 paper it is assumed that the likelihood does not decompose into independent  
161 individual components, but the individual likelihoods for the different data  
162 sets share (at least) some parameters.

163 The conditional independence structure of the global likelihood also has an  
 164 interesting interpretation within a Bayesian analysis. Without loss of generality  
 165 suppose there are two independent data sets, so that  $n = 2$ . The posterior  
 166 distribution can be expressed as,

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{x}) &\propto f_G(\mathbf{x}; \boldsymbol{\theta})p(\boldsymbol{\theta}) & (1) \\ &= f_1(\mathbf{x}_1; \boldsymbol{\theta})f_2(\mathbf{x}_2; \boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\quad (\text{since } \mathbf{x}_1 \text{ and } \mathbf{x}_2 \text{ are independent given model parameters, } \boldsymbol{\theta}) \\ &\propto f_1(\mathbf{x}_1; \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x}_2). \end{aligned}$$

167 In other words the posterior distribution of the parameters given all the  
 168 available data can be re-expressed as the posterior distribution of the param-  
 169 eters given data set  $\mathbf{x}_1$ , with an associated prior distribution equal to the  
 170 posterior distribution of the parameters, given data set  $\mathbf{x}_2$ . This observation  
 171 has practical model-fitting implications (see Section 3.2), and immediately  
 172 extends to a more general number of data sets,  $n > 2$ .

## 173 2.2 Dependent data

174 In most cases, individual data sources are not independent of each other. Con-  
 175 sequently, the joint likelihood of the dependent data sets,  $f_G(\mathbf{x}; \boldsymbol{\theta})$ , cannot be  
 176 simplified. In the presence of both dependent data sets  $(\mathbf{x}_1, \dots, \mathbf{x}_k)$  and inde-  
 177 pendent data sets  $(\mathbf{x}_{k+1}, \dots, \mathbf{x}_n)$ , the joint likelihood of all available data can  
 178 be expressed in the form,

$$179 \quad f_G(\mathbf{x}; \boldsymbol{\theta}) = f_g(\mathbf{x}_1, \dots, \mathbf{x}_k; \boldsymbol{\theta}) \prod_{i=k+1}^n f_i(\mathbf{x}_i; \boldsymbol{\theta}),$$

180



181 where  $f_g(\mathbf{x}_1, \dots, \mathbf{x}_k; \boldsymbol{\theta})$  denotes the joint likelihood over the dependent data  
182 sets. Dependent data may arise due to the same individuals being studied in  
183 the given data collection surveys and hence resulting data sets. For example,  
184 this occurs in the case of capture-recapture and tag-recovery data. The same  
185 individuals appear in both the capture-recapture and tag-recovery data (e.g.  
186 following individuals being marked they may be observed at future periods  
187 both alive and dead), and the joint capture-recapture-recovery likelihood needs  
188 to be considered [26–30]. Alternatively, and particularly for populations that  
189 are geographically closed, census data combined with marked individuals (such  
190 as capture-recapture data) will often involve the same individuals in both sets  
191 of survey data [31]. As a final dependent data example, data may be collected  
192 on multiple interacting species within an area (and possibly data from different  
193 monitoring schemes) [21–23].

194 In practice data sets may be assumed to be approximately (conditionally)  
195 independent within an analysis, even when it's not strictly the case, as this  
196 dramatically simplifies the likelihood expression i.e. the global likelihood can  
197 be decomposed into the product of the individual likelihoods for each data set.  
198 The impact of assuming independence between dependent data sets has been  
199 examined using simulation, see for example [32–34]. From their studies, which  
200 combined count data with demographic data, they concluded that the amount  
201 of information contained in the demographic data relative to the survey data  
202 influenced the magnitude of the effect of violating the independence assump-  
203 tion. In particular, the simplest and most immediate effect was an inflated  
204 level of precision on the model parameters; with biases also observed in some  
205 cases, albeit generally small. As described in [32] such explorations have typ-  
206 ically focused on specific types of data being combined and the outcomes are

207 often used as justification that dependence has little effect on parameter esti-  
208 mates. However, a more general exploration of this area is required, especially  
209 as the types of data being included in IPMs is rapidly expanding.

210 To avoid the issues of potential bias and/or over-confidence in the preci-  
211 sion of the parameter estimates that may arise due to the same individuals  
212 appearing in multiple data sets (so that the data are non-independent) one  
213 approach has been to partition the individuals into the different data sets,  
214 so that they only belong to a single data set. This removes the dependence  
215 as individuals are no longer common to multiple data sets, but reduces the  
216 amount of information in each individual data set due to the reduced sample  
217 size. For example, [35] consider this approach for combining constant effort  
218 (mist-netting) count data and capture-recapture data, where the same indi-  
219 viduals are recorded in both survey methods. The data were partitioned at  
220 the geographical site level, with approximately half the sites allocated to the  
221 constant effort data; and the other half for the capture-recapture data. Within  
222 their application, they concluded that the split data integrated analysis led to  
223 substantially improved precision of the parameter estimates compared to the  
224 analysis of only the constant effort data; and only marginally wider credible  
225 intervals compared to considering an integrated analysis assuming the data  
226 were independent. However, we note that in many situations, data may be  
227 limited, and the approach requires some knowledge relating to their overlap or  
228 dependence (e.g. individuals may be uniquely identifiable). In practice, there  
229 is a trade-off between the removal of potential dependence and the associated  
230 reduction of information with the data. A sensitivity analysis, using different  
231 data components may help gain some insight into this trade-off. Further, and  
232 in general, the robustness of the results to splitting the data can be assessed

233 by considering different data splits, which for [35] appeared to be minimal for  
 234 their application.

### 235 **3 Integrated Population Models**

236 Ecological time-series data of species abundance are a very common input  
 237 within IPMs, as they provide direct information on abundance as well as the  
 238 (indirect) demographic processes. State-space models provide a structured way  
 239 of describing such time series data and can be viewed as a special case of a wider  
 240 class of models known as hidden process models [18, 36–39]. These models can  
 241 be described via two separate processes: (i) the state process that describes  
 242 the evolution of the true underlying (unobserved) state-vector corresponding  
 243 to true abundance over time; and (ii) the observation equation that links the  
 244 elements of the state vector to the observed data at each time point. Given  
 245 the prominence of these data within IPMs, their general structure of them is  
 246 briefly described.

247 Suppose observed data correspond to a multivariate time series over dis-  
 248 crete time events,  $t = 1, \dots, T$ , which are denoted by  $\mathbf{y} = \{\mathbf{y}_t : t = 1, \dots, T\}$ ,  
 249 with each observation  $\mathbf{y}_t$  a vector containing  $K$  elements. The observed data  
 250 are related to an  $m \times 1$  vector,  $\boldsymbol{\alpha}_t$ , known as the state vector via the observation  
 251 equation,

$$252 \quad \mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t,$$

254 for  $t = 1, \dots, T$ , where  $\mathbf{Z}_t$  denotes an  $K \times m$  matrix and  $\boldsymbol{\epsilon}_t$  an  $K \times 1$  vector  
 255 corresponding to the observation error. In general, the elements of  $\boldsymbol{\alpha}_t$  are not  
 256 observable, but are assumed to be first-order Markovian, such that the state  
 257 equation can be expressed in the form,

258  
259

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t,$$

260 where  $\boldsymbol{\eta}_t$  denotes an  $m \times 1$  vector corresponding to the system process error.

261 For ecological applications the state vector,  $\boldsymbol{\alpha}_t$ , often relates to the number  
262 of individuals in particular age (and/or state) classes. The matrix,  $\mathbf{T}_t$ , governs  
263 the evolution of the state-vector from occasion  $t$  to occasion  $t + 1$ , and is typi-  
264 cally expressed as a function of demographic parameters, such as the survival  
265 probabilities, fecundity rates and/or migration rates. The formation of  $\mathbf{T}_t$  can  
266 be fairly straightforward for state vectors of small dimension, however it is  
267 often useful to decompose the formation of  $T$  into intermediate sub-processes,  
268 such as survival, ageing, recruitment etc [37, 40].

269

**Example:**

Following [19], consider a two-age class population. Let  $\alpha_{1,t}$  denote the number of individuals in their first-year of life and  $\alpha_{a,t}$  denote the number of individuals older than 1 year (i.e. “adults”) at time  $t$ . A possible state-equation for this population is given by

$$\begin{pmatrix} \alpha_1 \\ \alpha_a \end{pmatrix}_{t+1} = \begin{pmatrix} 0 & \rho\phi_1 \\ \phi_a & \phi_a \end{pmatrix}_t \begin{pmatrix} \alpha_1 \\ \alpha_a \end{pmatrix}_t + \boldsymbol{\eta}_{t+1},$$

for  $t = 1, \dots, T - 1$ , where  $\phi_{1,t}$  denotes the probability of first-year survival,  $\phi_{a,t}$  denotes the probability of adult survival and  $\rho_t$  denotes the productivity parameter at time  $t$ . The system process error (often referred to as demographic stochasticity [41]),  $\boldsymbol{\eta}_{t+1}$  will typically assumed to have mean zero, with some error structure, such as Poisson for age 1 and Binomial for those older than 1 year.

If only adult individuals are observed during a census count, the corresponding observation equation may be expressed as,

$$y_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_a \end{pmatrix}_t + \epsilon_t,$$

for  $t = 1, \dots, T$ , with  $\epsilon_t \sim N(0, \sigma^2)$ , and observation error parameter  $\sigma^2$ .

A closed form likelihood for state-space models is only available when specifying either (i) a linear and Gaussian model, or (ii) where the state vector is discrete-valued, leading to a hidden Markov model (HMM). See [39] for further discussion. The information contained within the temporal abundance data alone may be relatively weak, in terms of the demographic parameters, which may be strongly correlated and/or even confounded. Thus,

271

272

273

274

275

276

277 such abundance data are often integrated with additional forms of data that  
278 provide information on more of the demographic parameters. For example,  
279 capture-recapture-type data may be used to provide information on survival  
280 probabilities; or nest-record data for productivity rates. For such data, the  
281 associated likelihood functions may often be expressed in closed form. How-  
282 ever, combining these different forms of data within an integrated modelling  
283 approach lead to a number of additional challenges and modelling consider-  
284 ations. Four practical challenges associated with the application of IPMs to  
285 ecological data are summarised in Table 1. Existing work that begins to address  
286 some of these challenges are discussed in Sections 3.1-3.4 before discussing  
287 potential avenues for further research in Section 4.

288 [Table 1 about here.]

### 289 **3.1 Challenge 1: Model specification**

290 For any given data set, and specified ecological question or hypothesis, the  
291 first step in the data analysis pipeline involves specifying a statistical model to  
292 describe the observed data. This requires knowledge of both the data collec-  
293 tion process and associated ecological system being studied. This information  
294 permits the construction of the statistical model, given appropriate assump-  
295 tions, and the associated likelihood function to be derived [1, 42–44]. However,  
296 for IPMs, there are multiple observation processes, each associated with the  
297 different datasets, and typically either multiple system processes (which may  
298 interact with each other) and/or more complex system processes. Conse-  
299 quently, IPMs lead to additional considerations within the model specification.  
300 One such consideration has already been discussed in Section 2 in relation to  
301 the specification of the joint likelihood function of the data and whether the  
302 data sets can be regarded as independent, conditional on the associated model

303 parameters. However, due to the combining of the different data sets, further  
 304 care is needed in terms of the interpretation (or equivalence) of the parameters  
 305 associated with the system process(es).

306 For multiple data sets the relationship between the parameters associated  
 307 with each data set needs to be considered with some care. In practice, there  
 308 will be some parameters for which there is direct information from only one  
 309 data set; and other parameters for which there is direct information from two  
 310 (or more) data sets. For example, suppose that there are  $n = 2$  data sets. The  
 311 set of parameters can be decomposed as follows:  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{1,2}\}$ , where  $\boldsymbol{\theta}_j$   
 312 corresponds to the parameters uniquely associated with the model for data  
 313 component  $j = 1, 2$ ; and  $\boldsymbol{\theta}_{1,2}$  denotes the parameters for which there is direct  
 314 information contained in both data sets. For simplicity the parameters for  
 315 which there is direct information from multiple data sets (i.e.  $\boldsymbol{\theta}_{1,2}$  in this  
 316 example) are referred to as the *common* parameters across the given data sets.  
 317 For example, consider integrated count data with ring-recovery data ( $\mathbf{x}_1 =$   
 318 count data;  $\mathbf{x}_2 =$  ring-recovery data) with model parameters  $\boldsymbol{\theta}$  corresponding  
 319 to demographic survival probabilities,  $\boldsymbol{\phi}$ , and fecundity rates,  $\boldsymbol{\rho}$ ; and associated  
 320 observation process parameters corresponding to recovery rates,  $\boldsymbol{\lambda}$  (for the  
 321 ring-recovery data), and observation error variance,  $\sigma^2$  (for the count data).  
 322 Then,  $\boldsymbol{\theta}_1 = \{\boldsymbol{\rho}, \sigma^2\}$ ;  $\boldsymbol{\theta}_2 = \boldsymbol{\lambda}$  and  $\boldsymbol{\theta}_{1,2} = \boldsymbol{\phi}$  (see, for example, [19] for further  
 323 discussion in relation to this particular example). Note that for the special case  
 324 of conditionally independent data sets the joint likelihood (for  $n = 2$ ) can be  
 325 written as,

$$326 \quad f(\mathbf{x}; \boldsymbol{\theta}) = f_1(\mathbf{x}_1; \boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2})f_2(\mathbf{x}_2; \boldsymbol{\theta}_2, \boldsymbol{\theta}_{1,2}),$$

327  
 328 with the immediate extension for  $n \geq 3$  data sets.

329 In considering the general model construction and parameter specification  
330 there are two particular points to emphasise:

- 331 1. The parameters that are common to different data sets *must* have the same  
332 interpretation across these different data sets; and
- 333 2. Considering the joint likelihood of the integrated model may permit the  
334 estimation of previously confounded parameters.

335 The first point requires knowledge of the different data sets and associated  
336 modelling assumptions. In particular, parameters that appear to be common  
337 to multiple data sets may have slightly different (possibly nuanced) defini-  
338 tions. For example, consider two data sets which each provide information on  
339 the survival probabilities of the species of interest. However, there may still be  
340 assumed differences in relation to the interpretation of these survival param-  
341 eters. For example, *true* survival (i.e. non-mortality) as opposed to *apparent*  
342 survival allowing for other departures from the study site, such as permanent  
343 migration [30]; or where the survival probabilities are specified over different  
344 geographical locations and/or temporal periods [45]. In practice, it may be of  
345 ecological interest whether given parameter(s) are equal across the different  
346 data sets, providing ecological insight into the systems, leading to poten-  
347 tial model selection (or hypothesis testing) within the statistical analysis (see  
348 Section 3.3). Given the parameter is common across data sets, applying an  
349 integrated modelling approach typically leads to improved precision of the  
350 parameter, due to the increased information available on the parameter [45].

351 Note that, perhaps somewhat ironically, having “similar” but distinct  
352 parameters for the different data sets (as for true and apparent survival above),  
353 can be very useful in IPMs as this can lead to the estimation of parameters that  
354 are confounded when only considering the individual data sets. To illustrate  
355 this mathematically, return to the simple two data set example. Suppose the



356 model processes for data set,  $\mathbf{x}_1$ , are a function of a true survival probability  $S$ ,  
357 and for data set,  $\mathbf{x}_2$ , this is a function of apparent survival probability,  $\phi$ , which  
358 is confounded with permanent migration from the study site. Let  $\gamma$  denote the  
359 probability of permanently migrating from the study site, so that  $\phi = (1 - \gamma)S$ .  
360 Then  $S \in \boldsymbol{\theta}_{1,2}$  and  $\gamma \in \boldsymbol{\theta}_2$ , with  $\phi$  now a derived parameter, calculated as a  
361 function of these terms [15, 30, 46]. More generally, however, parameters that  
362 are confounded when considering a single data set, may be estimable when  
363 combined with additional, relevant information. For example, fledgling sur-  
364 vival and first year survival are confounded for ring-recovery data if rings are  
365 applied to chicks in the nest and rings are recorded at the (coarse) annual level.  
366 However, additional nest record data may provide direct information on the  
367 fledgling survival, which when combined with the ring-recovery data permits  
368 direct estimation of the first year (post-fledgling) survival [47]. Alternatively,  
369 count data is often focused on the number of (adult) breeding birds, often  
370 leading to the associated model with first year survival and productivity con-  
371 founded. An IPM, incorporating additional capture-recapture/ring-recovery  
372 data, permits the estimation of first-year survival, and hence the estimation of  
373 productivity rates [7, 19, 48]; or additional nest-record data provides data on  
374 productivity rate, and in turn the first-year survival rates [49].

### 375 **3.2 Challenge 2: Computational aspects**

376 To combine data sources together, as previously discussed, IPMs typically  
377 require a more indepth and/or complex model structure than the analy-  
378 sis of data sources individually. While each model component of an IPM  
379 taken independently may be easily fitted to data in a standard framework,  
380 their integration with abundance data expressed within a state-space model-  
381 ing framework often leads to additional computational challenges [39]. This

382 increased complexity leads to greater computational requirements in terms of  
383 algorithm complexity, run times, and memory load. For example, standard  
384 MCMC (data augmentation) techniques that are widely used in practice, can  
385 require days or even weeks to run IPMs on desktop computers using dedi-  
386 cated black-box software, such as JAGS [50]. The computational burden is  
387 such that it becomes difficult – if not impossible – to go through the usual  
388 modelling strategy of starting simple and adding complexity, or to fit several  
389 models and compare corresponding ecological hypotheses (see Section 3.3).  
390 Note that the most recent software like Nimble [51] allows choosing and cus-  
391 tomising MCMC algorithms which may help optimising time computation for  
392 specific components of IPMs. Here we describe two alternative strategies that  
393 are often applied to address the computational model-fitting challenges.

### 394 **Consideration of separate IPM components**

395 When building IPMs, for the case of conditionally independent data sets, it  
396 is common practice to go through building each data component separately  
397 before combining them [52]. This approach focuses on each individual com-  
398 ponent in turn and provides a natural approach to identifying the associated  
399 computational burden by identifying a possible bottleneck, and optimising the  
400 fitting of the corresponding component. A bottleneck most naturally occurs  
401 when the model is specified via unobserved (latent) variables, such as state-  
402 space models for abundance/count data (with unobserved population sizes,  
403 possibly over several (st)ages); multi-state capture-recapture models (such  
404 that states are unknown when an individual is unobserved); or individual  
405 heterogeneity model (with unobserved random effect terms).

406 More generally, several approaches have been developed to improve com-  
407 putational times for individual model components. This often involves imple-  
408 menting strategies that lead to faster evaluations of the likelihood function.

409 For example, [53] marginalise the likelihood, removing the latent states from  
410 the model, and hence decreases the number of parameters to be estimated.  
411 Similarly, for multi-state capture-recapture models [29, 54] marginalise the  
412 likelihood to provide closed form expressions via efficient sufficient statistics,  
413 further facilitating its efficient calculation. These marginalisation approaches  
414 lead to more complex (observed-data) likelihood functions, compared to the  
415 augmented (complete-data) likelihood with additional latent variables. The  
416 marginal likelihood can be maximised directly within a frequentist framework;  
417 or needs to be evaluated substantially fewer times when using a Bayesian  
418 Markov chain Monte Carlo (MCMC) implementation leading to faster compu-  
419 tational time and typically better mixing, compared to a latent variable data  
420 augmentation implementation.

421 Alternatively, approximate likelihood approaches may be applied to dif-  
422 ferent modelling components. For example, state-space models, using linear  
423 and/or Gaussian approximations for the system and observation processes,  
424 permits the use of the fast and efficient Kalman filter algorithm [55] to evalu-  
425 ate the likelihood function. Alternatively, (at least for system processes of low  
426 dimension), a coarse discretisation (or “binning” approach) may be applied  
427 to the system states, leading to an HMM approximation [56–58]. In partic-  
428 ular, [59, 60] used binning to deal with the large number of possible states  
429 for the state vector of population abundance, and showed that this numerical  
430 approximation performs well compared to the Kalman filter approximations  
431 and MCMC simulations. Further [60] proposed a semi-complete data likeli-  
432 hood approach [61] to improve computational efficiency of fitting more complex  
433 state-space models using a combined data augmentation and numerical inte-  
434 gration scheme. They also demonstrated improved computational efficiency

435 using a binning approach, with minimal impact on the estimation of the param-  
436 eters. Overall, dealing with an intractable likelihood for general state-space  
437 models limits computational model-fitting tools, and more research is needed  
438 in that direction [39, 62].

### 439 **Full IPM likelihood evaluation**

440 An alternative to considering the separate IPM components, is to optimise the  
441 evaluation of the full IPM likelihood. Note that this is, in general, necessary  
442 when the data are dependent, as the likelihood cannot be factored into the  
443 separate components, but can also be applied to the conditionally independent  
444 case. In general this process may also involve optimising the evaluation of  
445 each separate data component as well. One general approach for the likelihood  
446 specification is to formulate the IPM (possibly via a suitable approximation) as  
447 an HMM, with multiple observation processes, and potentially multiple system  
448 processes [63]. Formulating the IPM in this framework opens access to the  
449 available toolbox of efficient algorithms to fit HMMs that have been specifically  
450 developed to improve computational efficiency. See [64] for further discussion  
451 of pitfalls and opportunities of HMM approximation to general state-space  
452 population dynamics models.

453 Exploring another possibility, [65] proposed an efficient methodology for  
454 fitting IPMs in a Bayesian framework. In particular they exploited the inte-  
455 grated model structure to reduce the computational cost of the algorithm by  
456 reducing the number of times the likelihood needed to be evaluated. More  
457 specifically, a delayed acceptance approach was implemented, where the com-  
458 putationally intensive part of the likelihood, corresponding to the state-space  
459 model for the count data, was only evaluated if the proposed parameter value  
460 in the MCMC algorithm was evaluated to be “good” in terms of the fast data  
461 likelihood component relating to the capture-recapture-recovery-type data.

462 For data sets that are conditionally independent, a further step-wise pro-  
463 cess can be applied when considering the full IPM likelihood. Recall that the  
464 joint posterior distribution of the parameters can be expressed as the product  
465 of the likelihood of a single data set, say  $\mathbf{x}_1$ , with prior equal to the posterior  
466 distribution of the parameters given the other data sets (Equation 1). For this  
467 case, where the associated prior is of standard form, the model-fitting is sim-  
468 plified to consideration of data set  $\mathbf{x}_1$ , given the posterior distribution of the  
469 parameters given the other data sets. This approach has been adopted by, for  
470 example, [66] and [67] where  $\mathbf{x}_1$  corresponded to multi-state capture-recapture-  
471 recovery data and  $\mathbf{x}_2$  to multi-state radio-tagging data, with the common  
472 parameters between the data sets the survival and state-transition probabili-  
473 ties. However, in many cases, the posterior distribution  $\pi(\boldsymbol{\theta}|\mathbf{x}_2)$  may not be  
474 of standard distributional form. In this case the posterior distribution may  
475 be approximated, for example, by specifying (independent) distributions with  
476 parameters determined using a moment-matching approach with the posterior  
477 summary statistics.

### 478 **3.3 Challenge 3: Model assessment**

479 Assessment for ecological models, encompassing both relative goodness-of-fit  
480 using model selection/comparison strategies, and absolute goodness-of-fit are  
481 both well grounded with standard procedures available in a practitioner's  
482 toolkit. However, model assessment for IPMs, combining multiple data sets  
483 within the same model-fitting process, is relatively underdeveloped.

#### 484 **Model selection**

485 The potential model space for integrated models can be very large. For exam-  
486 ple, demographic parameters may be time or state dependent and/or depend  
487 upon individual or time-varying covariates. In addition, error structures within

488 a state-space model for longitudinal time series can vary over time and/or  
489 space. The combination of different parameter dependencies and/or error  
490 structures often leads to a large number of possible models that may be fit-  
491 ted to the data. Individually fitting all such possible models to data can be  
492 time consuming or even simply infeasible, in practice. This typically leads to  
493 the use of a step-wise search algorithm over the model space [45]. Due to the  
494 lack of robust model selection approaches for IPMs, *ad-hoc* step-wise model  
495 selection approaches are often applied separately to the different modelling  
496 components, either at the parameter level or data set level. For example, the  
497 structure of an IPM may be investigated step-wise using the data set(s) which  
498 contains the most information about a parameter (in terms of precision). This  
499 approach is applied by [19] when combining ring-recovery and count data, with  
500 the model selected using the ring-recovery component for the survival compo-  
501 nent (common to both data sets); with subsequent parameter dependence for  
502 the productivity for the count model determined, conditional on the survival  
503 dependence already selected. Implementing such search algorithm strategies,  
504 however, can lead to different results when compared to model selection con-  
505 ducted on the global, integrated, model and a suboptimal model being selected  
506 [68].

507 For some forms of data, such as capture-recapture-recovery-type data,  
508 model selection using standard information criteria (or adaptations of infor-  
509 mation criteria to account for small sample sizes or overdispersion) generally  
510 appears to work well [69]. However information criteria do not appear to  
511 perform as well for HMMs and/or state-space models, particularly where com-  
512 peting models differ in terms of the dimension of the (unobserved) state vector.  
513 For example, information criteria have been shown to overestimate the number

514 of states within HMMs [70]. Information criteria within the Bayesian frame-  
515 work have also been applied to IPMs, such as the DIC [14, 71] and WAIC [72].  
516 Although easy to compute, there has been no formal evaluation of the perfor-  
517 mance of these criteria for IPMs. Specific information criteria for state-space  
518 models have been proposed (see for example [73]), however these are compu-  
519 tationally intensive and their extension to state-space models when integrated  
520 with other models, as in an IPM, has not been investigated. As an alternative,  
521 [74] propose a step-wise approach to determine the dimension of the state vec-  
522 tor (representing the age-structure of the population), and hence number of  
523 states required in the state vector, within an IPM framework and provides a  
524 starting-point for further research in this area.

525 Within the Bayesian framework there are a number of standard model selec-  
526 tion tools (see [75] and [76] for general guidance) which can be implemented in  
527 an IPM setting. Models can be quantitatively compared via posterior model  
528 probabilities, or Bayes Factors [77]. There are typically two particular chal-  
529 lenges in their use relating to (i) their estimation; and (ii) their sensitivity to  
530 prior distributions specified on the model parameters. To estimate posterior  
531 model probabilities, reversible jump (RJ)MCMC has been applied in rela-  
532 tion to the dependence structure of parameters, such as time and/or covariate  
533 dependence [48, 76]. One particular attraction of RJMCMC is that only a sin-  
534 gle MCMC chain needs to be run, as the MCMC chain is able to move over  
535 the model space. However, the RJMCMC algorithm can be difficult to imple-  
536 ment in the general case and requires bespoke code, limiting its application in  
537 practice; although for the special case of variable selection, such as covariate  
538 selection for given demographic parameters, an indicator variable approach

539 can be applied [51, 76, 78]. Alternatively, for state-space models, [65] consid-  
540 ered an approach using sequential Monte Carlo samplers that permitted the  
541 calculation of posterior model probabilities across a set of different IPMs.

542 Approaching model selection separately for each component of an IPM is  
543 attractive for its simplicity since bespoke model selection approaches which  
544 have been optimised for a particular data type can be used. However, iden-  
545 tifying a model as being the best from a candidate model set does not mean  
546 that the model actually fits the data well, and for IPMs in particular, even if  
547 individual components of the IPM fit well this does not necessarily mean that  
548 the overall IPM fits well. Therefore, assessment of fit of the overall IPM (as  
549 well as assessment of the appropriateness of any modelling assumptions made)  
550 is an essential step in an IPM analysis.

### 551 **Absolute goodness-of-fit**

552 Absolute goodness-of-fit tests for ecological models can be particularly use-  
553 ful for understanding disparities between the simplifying statistical model  
554 for a given system and the associated dynamics of the population under  
555 study. Gaining insight into potential lack of fit, can provide ecological under-  
556 standing and lead to improved ecological models. However, techniques for  
557 assessing absolute goodness-of-fit typically differ across the types of data  
558 (and models) available. For example, absolute goodness-of-fit techniques for  
559 capture-recapture-recovery-type models are fairly advanced, including diag-  
560 nostic goodness-of-fit tests which link the detection of lack of fit with a  
561 determination of the likely biological underlying cause of the inadequacy of the  
562 model [79–81]. However, goodness-of-fit assessment for ecological state space  
563 models is substantially more limited. The appropriateness of the Gaussian  
564 assumptions made underlying the use of Kalman filter recursions can be inves-  
565 tigated through diagnostic checks of normality of the prediction errors arising



566 from the recursions, but are unlikely to help guide adaptation of the model to  
567 improve fit [19]. To date, where such goodness-of-fit tests do exist, they are  
568 considered separately for each type of data and not at the integrated level.

569 One approach, often referred to as posterior predictive checks, is to con-  
570 sider the evaluation of a discrepancy measure between the observed data and  
571 data simulated from the fitted model. For example, in a Bayesian framework,  
572 this leads to the idea of a Bayesian  $p$ -value, where multiple data sets are  
573 simulated from the posterior distribution of the model and compared to the  
574 observed data to see if these are “similar”. See [82] for further discussion, and  
575 additional Bayesian approaches. Similarly, in a frequentist setting, the idea  
576 of calibrated simulation has been proposed, which implements a parametric  
577 bootstrap to obtain simulated data sets which are subsequently compared to  
578 the observed data via a given discrepancy measure [83]. In general, Bayesian  
579  $p$ -values and calibrated simulation rely on the specified discrepancy measure,  
580 and conclusions may vary dependent on the specific measure used. Further,  
581 for different types of data sets, one may wish to consider different discrepancy  
582 measures, which leads to further challenge when combined within an integrated  
583 framework.

### 584 **3.4 Challenge 4: Forecasting**

585 Forecasting ecological population sizes or trends is of particular interest within  
586 conservation management and, driven by the ecological incentive, there exist  
587 tools for forecasting across wide-ranging time scales. For instance, short-term  
588 forecasting tools are typically used in the regular management and conser-  
589 vation of a species, for example, in the fisheries industry to regulate annual  
590 harvest capacity of waters [84, 85]. In contrast, long-term forecasting tools

591 are especially useful for prioritising vulnerable species for conservation inter-  
592 vention [86], evaluating the viability of a population of a reintroduced species  
593 [87, 88], and in projecting the minimum effort needed to locally eradicate an  
594 invasive species [2]. Despite the rapid growth in technological tools for mea-  
595 suring, monitoring and analysing species data, robust methods of ecological  
596 forecasting are still an underdeveloped area [84]. One of the biggest hindrances  
597 to forecasting the future trajectories of animal populations (regardless of time  
598 scale) is the limited amount of data available for precise inference on how  
599 environmental variables influence demographic parameters. High precision on  
600 estimates of covariate effects is critical for prediction, since forecast errors will  
601 magnify over time. Thus, when parameter precision is poor, forecasts will be  
602 rendered effectively useless ([89] demonstrate this point with climate predic-  
603 tions). However, by increasing the amount of data available for inference via  
604 integrated modelling techniques, it is possible to improve the precision of demo-  
605 graphic parameters and their relationships to covariates and thus alleviate  
606 some uncertainty in species' projections [3, 90].

607 Various time-series methods have been applied to ecological data to inform  
608 decision-making on short-term time scales (days to years), such as random-  
609 walk, autoregressive (AR-1) and moving average (MA-1) models [91–93].  
610 However, such standard time-series techniques often require large datasets,  
611 which may be available for example for finance and climate, but rare in ecolog-  
612 ical settings and particularly so in newly established research or conservation  
613 initiatives.

614 Long-term forecasting is often used to assess the possible fates of pop-  
615 ulations. Such forecasts underpin Population Viability Analysis (PVA), a  
616 procedure of estimating the probability that a species will persist for a cer-  
617 tain amount of time [94–96]. By coupling IPMs with PVA, predictions of

618 demographic parameters can be improved by specifically incorporating mul-  
619 tiple sources of uncertainty in a unified framework. This may be particularly  
620 useful when forecasting populations of multiple interacting species with dis-  
621 parate data types. For example, this has been used to identify components of  
622 multi-species population cycles and evaluate the efficacy of different manage-  
623 ment strategies such as assessing how removals of one species may affect the  
624 population viability of the other [3]. Other work that has used IPMs within  
625 a PVA include, but are not limited to, forecasting future salmon populations  
626 in response to fishery exploitation [85], and predicting the population effects  
627 after environmental trauma such as large oil-spills [97].

628 An emerging area in ecological forecasting couples population models with  
629 climate projections to forecast populations over a number of decades [98, 99].  
630 The main assumption behind such approaches is that species will respond to  
631 future climate conditions similarly to their responses in the past. This is par-  
632 ticularly useful for providing evidence for practitioners of how populations  
633 may respond to certain climate scenarios and enables long-term risk assess-  
634 ment of species status. For example, [100] provided projections of the monarch  
635 butterfly population based on a range of climate change scenarios. However,  
636 whilst these environmental-based techniques provide useful insights into the  
637 future, they are not without their issues [101]. As the projection duration  
638 increases so does uncertainty in the parameter estimates and forecasts, espe-  
639 cially if the future climate scenario is significantly different from the past and  
640 current [102]. Further, these forecasting methods simply involve iterating the  
641 process model for as long a time as needed past the observation period. How-  
642 ever, when a population is rapidly changing and/or available retrospective data  
643 sets contain relatively short time frames of data, methods of iterating forward  
644 may not be very meaningful since the parameter estimates may be estimated

645 from unstable or rapidly changing populations (for example, when reintroduc-  
646 ing a species). Thus there is a need for dynamic models that can incorporate  
647 statistical non-stationarity [103].

648 In general one of the main reasons for inaccurate forecasting is a lack of  
649 available data and of methods accounting for multiple sources of uncertainty.  
650 IPMs may alleviate these issues, thus improving forecasts, by pooling informa-  
651 tion across data sources. In other words, the precision of parameter estimates  
652 in relatively data-poor sources can be improved by borrowing information from  
653 richer data sources [85].

## 654 4 Discussion

655 IPMs provide a statistically robust approach for integrating multiple sources of  
656 data, making use of all available information. Understanding the dependence  
657 between the multiple different forms of data sets and associated relationships  
658 will continue to lie at the interface of ecological data science. Close collab-  
659 oration between ecologists and statisticians is essential in order to construct  
660 biologically meaningful and statistically robust models. Further, there are  
661 additional challenges in relation to interpreting meaningful results, including  
662 rigorous goodness-of-fit assessment, and use of the associated model outputs  
663 to inform conservation management, for example, via future predictions that  
664 include propagating the associated parameter uncertainties. Further, efficient  
665 computational algorithms and user-friendly software are critical for the mod-  
666 els to be widely applied in practice. The outstanding challenges of the different  
667 aspects of applying and interpreting IPMs relating to: model specification;  
668 computational aspects; model assessment; and forecasting are discussed before  
669 final concluding remarks are provided.

**Model specification**

A range of challenges arise in the different components of the model specification. In particular, the form of available data is continually evolving, particularly with advances in technology [104]. For example, citizen science data collection continues to grow in popularity [105, 106] and eDNA data collection is increasingly being used due to its capability to detect multiple species from water or air samples [107]. Remote sensing technology such as drones are providing finer-scale aerial survey data of animals [108], satellite earth observation data over fine scales of 30-50cm are becoming available across larger geographical areas and acoustic recording technology is enabling monitoring of elusive marine species [109], from which machine learning techniques can provide population count estimates [110]. The associated statistical models and tools are being developed for these new forms of data, but additional issues arise with incorporating such data into IPMs due to differences, for example, in relation to quality, quantity and scale [111, 112]. New challenges arise relating to how data of potentially very different geographical scales may be integrated. For example, where a large data set of “poor” quality (i.e., low information content) may be combined with small data set(s) of “high” quality within a robust and rigorous framework; and how the relative information can be computed across varying scales and levels of missing/incomplete data. Simple evaluation of the relative information in component data sets within an IPM has been proposed through evaluation of generalised variances [113]. However there is a need for exploration of whether likelihood components could be weighted to reflect the varying quality of the available data.

Considering multi-species predator-prey models provides new insight into the dynamics of the wider ecosystem, as opposed to an individual single-species study [22, 23]. However, it is necessary for equilibrium conditions to be assessed

697 for valid interpretation of output from such models [114]. Understanding how  
698 more complex models such as these can be incorporated within the general  
699 IPM framework will provide greater flexibility and potential for such data  
700 integration, with direct implications for wildlife management and conservation.

701 Further, as different forms of data are collected, and combined within  
702 IPMs, additional statistical models will be constructed. Even for relatively  
703 well-studied types of data and associated “standard” biologically sensible mod-  
704 els, due to the specific observed data, this may lead to identifiability issues  
705 and confounded parameters that cannot be reliably estimated. Such issues  
706 are increasingly likely to arise with increasingly complex data. Analytic tools  
707 exist to determine if the model is parameter redundant and if so the estimable  
708 parameters of a model [115]. However these typically only consider each indi-  
709 vidual (independent) model component associated with a particular type of  
710 data. Identifying parameter redundancy increases in complexity as the mul-  
711 tiple types of data are combined within an IPM, so that analytic techniques  
712 may quickly become computationally infeasible and unable to scale to complex  
713 models. Formal determination of estimable parameters have been identified  
714 for the combination of count and ring-recovery data [9], but more practical  
715 exploration of identifiability, potentially with the use of numerical techniques  
716 [78, 116] may be required in practice.

## 717 **Computational aspects**

718 Modern model-fitting to data typically reduces to a computational problem:  
719 either a numerical optimisation problem to obtain the MLE of the parameters;  
720 or an MCMC sampling problem, in order to be able to obtain an estimate of the  
721 posterior distribution of interest. Additional issues arise, for example, when an  
722 associated likelihood component is not available in closed form (as for general

723 state-space models or in the presence of missing data), or where the likeli-  
724 hood is computationally expensive to evaluate. Such computational challenges  
725 often lead to specialised algorithms being developed. However, in general, this  
726 requires additional coding experience and is bespoke to the particular prob-  
727 lem being addressed. More general and easy-to-use computational solutions  
728 that can be applied to a wide suite of integrated models and applications are  
729 required for wider dissemination and impact.

730 Approximate likelihood approaches are an interesting alternative that  
731 can alleviate the computational expense of IPM components. For exam-  
732 ple, constructing capture-recapture-recovery type likelihoods in terms of  
733 maximum-likelihood estimates and corresponding variance-covariance matrix  
734 using a multivariate normal approximation has been shown to work well  
735 in an IPM framework [117]. This provides a potential mechanism for spe-  
736 cialised (optimised) computer packages to be applied to specific data types,  
737 which can then be combined with other data sets within an IPM frame-  
738 work. This has been extended further for an IPM combining census data  
739 with capture-recapture-type data. In this case a further approximation was  
740 made, assuming independence between the estimated MLEs of the parameters  
741 from the capture-recapture data (i.e. assuming a diagonal variance-covariace  
742 matrix) within the multivariate model approximation [45]. This suggests that  
743 published model estimates may be more widely used within an integrated  
744 framework, to reduce the computational burden and widen the potential appli-  
745 cation of IPMs, and the Bayesian framework through informative priors seems  
746 like a natural solution to do so. However, the application of such approxima-  
747 tions to date has been relatively limited, with further investigation required  
748 to more fully explore their potential and also associated limitations (see also  
749 additional model assessment challenges).

750 As data sets increase in size and/or models become more complex addi-  
751 tional efficient computational optimisation or sampling algorithms may be  
752 required. For example, by making use of structural properties of the likeli-  
753 hood function that may lead to improved optimisation algorithms; or reducing  
754 the size of the data set by considering subsamples of the data and correct-  
755 ing the associated posterior estimates [118]. Further, it may be possible to  
756 take advantage of the flexibility of Bayesian black-box software such as Stan  
757 [119] and Nimble [120] which permit customisation of the algorithms. This  
758 approach requires more intricate knowledge of the software, but the effort  
759 can substantially improve the model-fitting process [121, 122] and is usually  
760 more accessible than alternatives such as C(++) [123]. Understanding the rela-  
761 tionship between the different components of an IPM may suggest potential  
762 techniques for efficient model-fitting [65]. Alternatively, general likelihood-free  
763 approaches have been developed for fitting complex models, such as approx-  
764 imate Bayesian computation (ABC; [124]) and synthetic likelihood [125].  
765 Exploring how such techniques may be applied to IPMs, is a potentially  
766 interesting avenue of future research.

## 767 **Model assessment**

768 Various model selection approaches have been implemented for IPM analyses,  
769 including the use of standard information criteria [45], and model struc-  
770 ture being determined by the most informative data set [19]. However, these  
771 approaches are limited since they have only compared models with the same  
772 length of state vector. [74] proposed an approach for selecting the appropriate  
773 age-structure for given model parameters in an IPM, which requires compari-  
774 son of models with state vectors of different dimensions. However, this study



775 was limited to selecting only age-structure and so extending the ideas to addi-  
776 tional dependence structures such as time, state and/or covariate dependence  
777 is as yet largely untested.

778 Further, simultaneously considering the multiple types of dependencies  
779 (for example, age, time, covariate etc.) on the different model parameters is  
780 an additional challenge, both increasing the number of combinations of pos-  
781 sible models that may be considered and also potentially the complexity of  
782 the models under consideration. Additionally, the performance of standard  
783 statistical approaches have not been fully explored for IPMs. This suggests  
784 potential avenues, including, for example, investigating regularisation methods  
785 for model selection (such as Lasso), and the use of weakly informative priors  
786 [126]. Similarly, for absolute goodness-of-fit assessments, the calibrated simu-  
787 lation approach discussed in Section 3.3 may be applied for detecting a lack  
788 of fit within IPMs. [127] investigated the calibrated simulation approach and  
789 existing diagnostics goodness-of-fit tests for capture-recapture data to detect  
790 specific departures from the fitted model and the diagnostic tests performed  
791 well when under scenarios of substantial departure from model assumptions  
792 and large sample size (e.g. density-dependence, immigration, capture hetero-  
793 geneity). However, where data are more sparse there was less power to detect  
794 mis-specified IPMs.

795 More generally, many ecological models, including longitudinal counts over  
796 time and demographic data such as capture-recapture-recovery-type data, can  
797 be expressed within an HMM framework [18]. Such models naturally extend to  
798 IPMs as demonstrated by [59]. Thus, model assessment techniques developed  
799 for HMMs may be considered more generally for such IPMs. In particular,  
800 [70] investigated different techniques for determining the number of latent

801 states within HMMs which could assist model selection for IPMs, in rela-  
802 tion to determining the age-dependence structures, for example. Alternatively,  
803 [128] developed a diagnostic goodness-of-fit test to determine whether the pro-  
804 posed latent structure in an HMM for partially-observed capture-recapture  
805 data was appropriate based on the observed data. The specification of com-  
806 ponents of IPMs all within the related state-space modelling framework also  
807 potentially suggests the use of forecast variance for goodness-of-fit assessment  
808 [129]. Developing these approaches, and in particular extending them to mul-  
809 tiple observation and/or system processes, provide future directions for the  
810 associated model assessment challenges of IPMs.

811 Approximate likelihoods for different model components of IPMs may be  
812 used to deal with the computational challenges, or where the raw data may  
813 not be available. However, when using estimates and their associated variance-  
814 covariance matrix to approximate the likelihood function as a component  
815 within an IPM it is not possible to change the structure of the parameters  
816 within that approximate likelihood function. Therefore, the potential model  
817 space is restricted due to the use of the approximate likelihood and hence it  
818 would seem sensible that such a restriction should potentially be penalised for  
819 within the model selection of an IPM. However, no formal evaluation has yet  
820 been conducted to address what penalties should be imposed.

821 New statistical developments for model assessment of IPMs need to be prac-  
822 tical, in terms of feasibility when the computational time for fitting an IPM is  
823 taken into account, and accessible to the wide user-community of IPMs. Thus,  
824 any new approaches developed need to be compatible with software that indi-  
825 viduals are using to fit IPMs and users need to be aware of potential limitations  
826 of what the methods can be used to diagnose. Additionally, considerable effort  
827 needs to be made toward disseminating research and encouraging uptake from  
828 wider audiences.

**Forecasting**

One major avenue of future research within forecasting is the quantification and reduction of prediction uncertainty. Failure to account for uncertainty when making decisions in ecology can lead to poor management and policy decisions. In short-term forecasts, reduction in uncertainty may be possible by iteratively updating forecasts in light of new data by gaining feedback, assessing effectiveness, and adapting models [84, 130, 131]. Long-term forecasting tools which use climatic data to predict abundance [100, 132] can experience a non-linear increase in uncertainty as the projection duration increases and their predictive skill can often vary because of the complex interactions between climate and population dynamics [133]. By decomposing the sources of uncertainty, [98] determined that the largest contributor was sampling variance. However, this can be easily reduced through larger sample sizes, or combining data sources i.e. through the use of an IPM. In addition, [101] suggests that parameter uncertainty can be reduced, over the near and long term, by collecting targeted data to better understand mechanistic links. Another possibility, which is useful when resources are limited, is to optimise sampling design by investigating the cost-benefit of certain data collection methods i.e. assessing whether the benefit of using more expensive monitoring methods are worth their possible reduction in uncertainty. Currently, there exists literature on optimising sampling design of specific data types such capture-recapture [134] and occupancy studies [135]. However, optimising sampling design in studies where multiple data types are integrated is a relatively unexplored area.

For multi-species systems, ensemble ecosystem modelling (EEM) [136, 137] provides a quantitative method for forecasting abundances in the future. EEM integrates species interaction networks and simulations of population models using the Lotka-Volterra equations as a standard predator-prey model. This

856 technique is specifically designed for predicting the abundance of interact-  
857 ing species after a predator reintroduction and is useful for assessing whether  
858 there will likely be any significant change in species abundance between  
859 pre- and post-reintroduction estimates. Whilst this allows for assessing large  
860 scale species networks, it can be computationally challenging when a network  
861 exceeds 10 species. Long-term forecasting EEM can typically only provide sug-  
862 gestions of possible scenarios and future states of system with associated risk  
863 and it does so with uncertainty, however EEM presents these uncertainties in  
864 a systematic way making it easier for end-users to make decisions [138]. One  
865 interesting possibility of future research could be to use the techniques of EEM  
866 within an IPM framework. For example, by using the species interaction net-  
867 works of EEM within an IPM to forecast populations in multi-species systems.  
868 Conversely, it would also be worthwhile investigating whether the use of IPMs  
869 within an EEM framework helps to improve estimates.

870 Finally, emerging work combining integrated population models with inte-  
871 gral projection models (referred to as IPM<sup>2</sup>) allows individual heterogeneity in  
872 demographic rates to be included within an IPM [139]. This improves forecast-  
873 ing accuracy by allowing subtle individual-level mechanics to drive population  
874 dynamics. Further development and investigation of such approaches provide  
875 interesting avenues of research in this area.

## 876 **Conclusion**

877 With the recent advances in data collection technology it is now possible  
878 to collect data at a range of spatial and/or temporal scales as well as from  
879 individual-based data collection towards community-level data collection. The  
880 IPM framework provides an adaptable and flexible approach that can accom-  
881 modate the different scales and upscale to provide a community-level statistical  
882 modelling approach. Overcoming the different statistical challenges for IPMs

883 presented within this paper will ensure that appropriate statistical methods are  
884 available for extracting intricate level information from the available data sets.  
885 As data collection technology and ecological theory continues to evolve, it is  
886 essential that the associated statistical developments keep pace and, crucially,  
887 are made accessible to a wide range of users. Raising awareness and utility of  
888 such tools will permit rigorous data-driven conservation decision-making.

889 **Acknowledgments.** FF and RM were supported by EPSRC grant  
890 EP/S020470/1. RK was supported by the Leverhulme research fellowship RF-  
891 2019-299. EFZ was supported by the U.S. National Science Foundation grant  
892 DBI-1954406. OG was supported by the French National Research Agency  
893 (grant ANR-16-CE02-0007). For the purpose of open access, the author has  
894 applied a Creative Commons Attribution (CC BY) licence to any Author  
895 Accepted Manuscript version arising from this submission.

896 **Conflict of Interest Statement.** On behalf of all authors, the correspond-  
897 ing author states that there is no conflict of interest.

## 898 References

- 899 [1] Williams, B., Nichols, J., Conroy, M.: Analysis and Management of  
900 Animal Populations. Academic Press, Cambridge, Massachusetts (2002)
- 901 [2] Jardine, S.L., Sanchirico, J.N.: Estimating the cost of invasive species  
902 control. *Journal of Environmental Economics and Management* **87**, 242–  
903 257 (2018). <https://doi.org/10.1016/j.jeem.2017.07.004>
- 904 [3] Saunders, S.P., Cuthbert, F.J., Zipkin, E.F.: Evaluating population  
905 viability and efficacy of conservation management using integrated pop-  
906 ulation models. *Journal of Applied Ecology* **55**, 1380–1392 (2018). <https://doi.org/10.1111/1365-2664.13080>
- 907

- 908 [4] DiRenzo, G.V., Zipkin, E.F., Grant, E.H.C., Royle, J.A., Longo,  
909 A.V., Zamudio, K.R., Lips, K.R.: Eco-evolutionary rescue promotes  
910 host–pathogen coexistence. *Ecological Applications* **28**, 1948–1962  
911 (2018)
- 912 [5] Royle, J.A., Dorazio, R.M. (eds.): *Hierarchical Modeling and Inference*  
913 *in Ecology*, p. . Academic Press, San Diego (2009). [https://doi.org/](https://doi.org/10.1016/B978-0-12-374097-7.50001-5)  
914 [10.1016/B978-0-12-374097-7.50001-5](https://doi.org/10.1016/B978-0-12-374097-7.50001-5). [https://www.sciencedirect.com/](https://www.sciencedirect.com/science/article/pii/B9780123740977500015)  
915 [science/article/pii/B9780123740977500015](https://www.sciencedirect.com/science/article/pii/B9780123740977500015)
- 916 [6] Kery, M., Royle, A.J.: *Applied Hierarchical Modeling in Ecology: Anal-*  
917 *ysis of Distribution, Abundance and Species Richness in R and BUGS*  
918 *vol. 2*. Academic Press, Cambridge, Massachusetts
- 919 [7] Brooks, S.P., King, R., Morgan, B.J.T.: A Bayesian approach to combin-  
920 ing animal abundance and demographic data. *Animal Biodiversity and*  
921 *Conservation* **27**, 525–529 (2004)
- 922 [8] Schaub, M., Abadi, F.: Integrated population models: a novel analy-  
923 sis framework for deeper insights into population dynamics. *Journal of*  
924 *Ornithology* **152**, 227–237 (2011)
- 925 [9] Cole, D.J., McCrea, R.S.: Parameter redundancy in discrete state-space  
926 and integrated models. *Biometrical Journal* **58**, 1071–1090 (2016)
- 927 [10] Dorazio, R.M.: Accounting for imperfect detection and survey bias in sta-  
928 tistical analysis of presence-only data. *Global Ecology and Biogeography*  
929 **23**, 1472–1484 (2014). <https://doi.org/10.1111/geb.12216>
- 930 [11] Zipkin, E., Zylstra, E., Wright, A., Saunders, S., Finley, A., Dietze, M.,

- 931 Itter, M., Tingley, M.: Addressing data integration challenges to link eco-  
932 logical processes across scales. *Frontiers in Ecology and the Environment*  
933 **19**, 30–38 (2021). <https://doi.org/10.1002/fee.2290>
- 934 [12] Maunder, M.N., Punt, A.E.: A review of integrated analysis in fisheries  
935 stock assessment. *Fisheries Research* **142**, 61–74 (2013). [https://doi.org/](https://doi.org/10.1016/j.fishres.2012.07.025)  
936 [10.1016/j.fishres.2012.07.025](https://doi.org/10.1016/j.fishres.2012.07.025)
- 937 [13] Fournier, D., Archibald, C.P.: A general theory for analyzing catch at age  
938 data. *Canadian Journal of Fisheries and Aquatic Sciences* **39**, 1195–1207  
939 (1982). <https://doi.org/10.1139/f82-157>
- 940 [14] Schaub, M., Gimenez, O., Sierro, A., Arlettaz, R.: Use of integrated  
941 modeling to enhance estimates of population dynamics obtained from  
942 limited data. *Conservation Biology* **21**, 945–955 (2007)
- 943 [15] Reynolds, T., King, R., Harwood, J., Frederiksen, M., Wanless, S.,  
944 Harris, M.: Integrated data analyses in the presence of emigration  
945 and tag-loss. *Journal of Agricultural, Biological, and Environmental*  
946 *Statistics* **14**, 411–431 (2009)
- 947 [16] Saunders, S.P., Farr, M.T., Wright, A.D., Bahlai, C.A., Jr., J.W.R.,  
948 Rossman, S., Sussman, A.S., Arnold, T.W., Zipkin, E.F.: Disentan-  
949 gling data discrepancies with integrated population models. *Ecology* **100**  
950 (2019)
- 951 [17] Schaub, M., Fletcher, D.: Estimating immigration using a Bayesian  
952 integrated population model: choice of parameterization and priors.  
953 *Environmental and Ecological Statistics* **22**, 535–549 (2015)
- 954 [18] King, R.: Statistical ecology. *Annual Review of Statistics and Its*

- 955 Application **1**, 401–426 (2014)
- 956 [19] Besbeas, P., Freeman, S.N., Morgan, B.J.T., Catchpole, E.A.: Integrating  
957 mark-recapture-recovery and census data to estimate animal abundance  
958 and demographic parameters. *Biometrics* **58**, 540–547 (2002)
- 959 [20] Zhao, Q., Heath-Acre, K., Collins, D., Conway, W., Weegman, M.D.:  
960 Integrated population modelling reveals potential drivers of demography  
961 from partially aligned data: a case study of snowy plover declines under  
962 human stressors. *PeerJ*, 9–12475 (2021)
- 963 [21] Lahoz-Monfort, J.J., Harris, M.P., Wanless, S., Freeman, S.N., Mor-  
964 gan, B.J.T.: Bringing it all together: Multi-species integrated population  
965 modelling of a breeding community. *Journal of Biological, Agricultural  
966 and Environmental Statistics* **22**, 140–160 (2017)
- 967 [22] Barraquand, F., Gimenez, O.: Integrating multiple data sources to fit  
968 matrix population models for interacting species. *Ecological Modelling*  
969 **411**, 108713 (2019). <https://doi.org/10.1016/j.ecolmodel.2019.06.001>
- 970 [23] Quérroué, M., Barbraud, C., Barraquand, F., Turek, D., Delord, K.,  
971 Pacoureaux, N., Gimenez, O.: Multispecies integrated population model  
972 reveals bottom-up dynamics in a seabird predator–prey system. *Ecolog-  
973 ical Monographs* **91**, 01459 (2021). <https://doi.org/10.1002/ecm.1459>
- 974 [24] Besbeas, P., Freeman, S.N., Morgan, B.J.T.: The potential of integrated  
975 population modelling. *Australian & New Zealand Journal of Statistics*  
976 **47**, 35–48 (2005). <https://doi.org/10.1111/j.1467-842X.2005.00370.x>



- 977 [25] Zipkin, E.F., Saunders, S.P.: Synthesizing multiple data types for biolog-  
978 ical conservation using integrated population models. *Biological Conser-*  
979 *vation* **217**, 240–250 (2018). [https://doi.org/10.1016/j.biocon.2017.10.](https://doi.org/10.1016/j.biocon.2017.10.017)  
980 [017](https://doi.org/10.1016/j.biocon.2017.10.017)
- 981 [26] Barker, R.J.: Joint modeling of live-recapture, tag-resight, and tag-  
982 recovery data. *Biometrics* **53**, 666–677 (1997)
- 983 [27] Catchpole, E.A., Freeman, S.N., Morgan, B.J.T., Harris, M.P.: Inte-  
984 grated recovery/recapture data analysis. *Biometrics* **54**, 33–46 (1998)
- 985 [28] King, R., Brooks, S.P.: Model selection for integrated recovery/recapture  
986 data. *Biometrics* **58**, 841–851 (2002)
- 987 [29] King, R., McCrea, R.S.: A generalised likelihood framework for partially  
988 observed capture–recapture–recovery models. *Statistical Methodology*  
989 **17**, 30–45 (2014). <https://doi.org/10.1016/j.stamet.2013.07.004>
- 990 [30] Sarzo, B., King, R., Conesa, D., Hentati-Sundberg, J.: Correcting bias in  
991 survival probabilities for partially monitored populations via integrated  
992 models. *Journal of Agricultural, Biological and Environmental Statistics*  
993 **26**, 200–219 (2021). <https://doi.org/10.1007/s13253-020-00423-1>
- 994 [31] Tavecchia, G., Besbeas, P., Coulson, T., Morgan, B.J.T., Clutton-Brock,  
995 T.H.: Estimating population size and hidden demographic parameters  
996 with state-space modeling. *American Naturalist* **173**, 722–733 (2009)
- 997 [32] Besbeas, P., Borysiewicz, R.S., Morgan, B.J.T.: Completing the ecologi-  
998 cal jigsaw. In: Thomson, D., Cooch, E.G., Conroy, M.J. (eds.) *Modeling*  
999 *Demographic Processes in Marked Populations* vol. 3, pp. 513–540.  
1000 *Environmental and Ecological Statistics*, Boston, MA (2008)

- 1001 [33] Abadi, F., Gimenez, O., Aslettaz, R., Schaub, M.: An assessment  
1002 of integrated population models: bias, accuracy, and violation of the  
1003 assumption of independence. *Ecology* **91**, 7–14 (2010)
- 1004 [34] Weegman, M.D., Arnold, T.W., Clark, R.G., Schaub, M.: Partial and  
1005 complete dependency among data sets has minimal consequence on esti-  
1006 mates from integrated population models. *Ecological Applications* **31**,  
1007 2258 (2021)
- 1008 [35] Cave, V.M., King, R., Freeman, S.N.: Integrated population model from  
1009 constant effort bird-ringing data. *Journal of Agricultural, Biological, and*  
1010 *Environmental Statistics* **15**, 119–137 (2010)
- 1011 [36] Newman, K.B., Buckland, S.T., Lindley, S.T., Thomas, L., Fernandez,  
1012 C.: Hidden process models for animal population dynamics. *Ecological*  
1013 *Applications* **16**, 74–86 (2006)
- 1014 [37] Newman, K.B., Buckland, S.T., Morgan, B.J.T., King, R., Borchers,  
1015 D.L., Cole, D.J., Besbeas, P., Gimenez, O., Thomas, L.: *Modelling Pop-*  
1016 *ulation Dynamics: Model Formulation, Fitting and Assessment Using*  
1017 *State-space Methods*. Springer, New York (2014)
- 1018 [38] Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R.,  
1019 King, A.A., Leos-Barajas, V., Mills Flemming, J., Nielsen, A., Petris, G.,  
1020 Thomas, L.: A guide to state–space modeling of ecological time series.  
1021 *Ecological Monographs* **91**, 1–38 (2021)
- 1022 [39] Newman, K., King, R., Elvira, V., de Valpine, P., McCrea, R.S., Mor-  
1023 gan, B.J.T.: State-space models for ecological time series data: Practical  
1024 model-fitting. *Methods in Ecology and Evolution* (2022). In press

- 1025 [40] Buckland, S.T., Newman, K.B., Thomas, L., Koesters, N.B.: State-  
1026 space models for the dynamics of wild animal populations. *Ecological*  
1027 *Modelling* **171**, 157–175 (2004)
- 1028 [41] Caswell, H.: The formal demography of kinship ii: Multistate models,  
1029 parity, and sibship. *Demographic Research* **42**, 1097–1144 (2020)
- 1030 [42] Hilborn, R., Mangel, M.: *The Ecological Detective: Confronting Models*  
1031 *with Data*. Princeton University Press, New Jersey (1997)
- 1032 [43] Clark, J.S.: *Models for Ecological Data: An Introduction*. Princeton  
1033 University Press, New Jersey (2007)
- 1034 [44] Bolker, B.M.: *Ecological Models and Data in R*. Princeton University  
1035 Press, New Jersey (2008)
- 1036 [45] McCrea, R.S., Morgan, B.J.T., Gimenez, O., Besbeas, P., Bregnballe,  
1037 T., Lebreton, J.-D.: Multi-site integrated population modelling. *Journal of Biological, Agricultural and Environmental Statistics* **15**, 539–561  
1038 (2010)  
1039
- 1040 [46] Burnham, K.P.: A theory for combined analysis of ring recovery and  
1041 recapture data. In: Lebreton, J.D., North, P.M. (eds.) *Marked Individ-*  
1042 *uals in the Study of Bird Population*, pp. 199–213. Birkhäuser Verlag  
1043 Basel, Switzerland (1993)
- 1044 [47] Thomson, D.L., Baillie, S.R., Peach, W.J.: A method for studying post-  
1045 fledging survival rates using data from ringing recoveries. *Bird Study*  
1046 **46**(sup1), 104–111 (1999) [https://arxiv.org/abs/https://doi.org/10.](https://arxiv.org/abs/https://doi.org/10.1080/00063659909477237)  
1047 [1080/00063659909477237](https://doi.org/10.1080/00063659909477237). <https://doi.org/10.1080/00063659909477237>

- 1048 [48] King, R., Brooks, S., Mazzetta, C., Freeman, S., Morgan, B.: Identifying  
1049 and diagnosing population declines: A Bayesian assessment of lapwings  
1050 in the UK. *Journal of the Royal Statistical Society Series C* **57**, 609–632  
1051 (2008). <https://doi.org/10.1111/j.1467-9876.2008.00633.x>
- 1052 [49] Robinson, R., Baillie, S., King, R.: Population processes in European  
1053 Blackbirds *Turdus merula*: A state-space approach. *Journal of Ornithol-*  
1054 *ogy* **152**, 419–433 (2012). <https://doi.org/10.1007/s10336-010-0612-y>
- 1055 [50] Plummer, M.: JAGS: A program for analysis of Bayesian graphical  
1056 models using Gibbs sampling. In: *Proceedings of the 3rd International*  
1057 *Workshop on Distributed Statistical Computing*, vol. 124, pp. 1–9 (2003)
- 1058 [51] de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Lang,  
1059 D., Bodik, R.: Programming with models: Writing statistical algorithms  
1060 for general model structures with NIMBLE. *Journal of Computational*  
1061 *and Graphical Statistics* **26**, 403–413 (2017)
- 1062 [52] Schaub, M., Kery, M.: *Integrated Population Models*. Elsevier, Cam-  
1063 bridge, Massachusetts (2021)
- 1064 [53] Turek, D., de Valpine, P., Paciorek, C.J.: Efficient Markov chain Monte  
1065 Carlo sampling for hierarchical hidden Markov models. *Environ Ecol*  
1066 *Stat* **23**, 549–564 (2016)
- 1067 [54] King, R., Brooks, S.P.: Closed-form likelihoods for Arnason–Schwarz  
1068 models. *Biometrika* **90**, 435–444 (2002)
- 1069 [55] Kalman, R.E.: A new approach to linear filtering and prediction prob-  
1070 lems. *Journal of Basic Engineering, Transactions ASMA, Series D* **82**,  
1071 35–45 (1960)

- 1072 [56] Kitagawa, G.: Non-Gaussian state-space modeling of nonstationary time  
1073 series. *Journal of the American Statistical Association* **82**, 1032–1041  
1074 (1987)
- 1075 [57] Langrock, R.: Some applications of nonlinear and non-Gaussian state-  
1076 space modelling by means of hidden Markov models. *Journal of Applied*  
1077 *Statistics* **38**, 2955–2970 (2011)
- 1078 [58] Langrock, R., King, R.: Maximum likelihood estimation of mark-  
1079 recapture–recovery models in the presence of continuous covariates.  
1080 *Annals of Applied Statistics* **7**, 1709–1732 (2013)
- 1081 [59] Besbeas, P., Morgan, B.J.T.: Exact inference for integrated population  
1082 modelling. *Biometrics* **75**, 475–484 (2019)
- 1083 [60] Borowska, A., King, R.: Semi-complete data augmentation for effi-  
1084 cient state-space model fitting. *Journal of Computational and Graphical*  
1085 *Statistics* (2022). In press
- 1086 [61] King, R., McClintock, B., Kidney, D., Borchers, D.L.: Abundance estima-  
1087 tion using a semi-complete data likelihood approach. *Annals of Applied*  
1088 *Statistics* **10**, 264–285 (2016)
- 1089 [62] Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis,  
1090 M.A., Jonsen, I.D., Flemming, J.M.: State-space models’ dirty lit-  
1091 tle secrets: even simple linear Gaussian models can have estimation  
1092 problems. *Scientific Reports* **6**, 1–10 (2016)
- 1093 [63] Beliveau, A.: Data integration methods for studying animal population  
1094 dynamics. PhD thesis, Simon Fraser University (2016)

- 1095 [64] Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T.,  
1096 McClintock, B.T.: Hidden markov models: Pitfalls and opportunities in  
1097 ecology. *Methods in Ecology and Evolution* (2022). In press
- 1098 [65] Finke, A., King, R., Beskos, A., Dellaportas, P.: Efficient sequen-  
1099 tial monte carlo algorithms for integrated population models. *Journal*  
1100 *of Agricultural, Biological and Environmental Statistics* **24**, 204–224  
1101 (2017). <https://doi.org/10.1007/s13253-018-00349-9>
- 1102 [66] Dupuis, J., Badia, J., Maublanc, M.-L., Bon, R.: Survival and spatial  
1103 fidelity of mouflon (*ovis gmelini*): A bayesian analysis of an age-  
1104 dependent capture-recapture model. *Journal of Agricultural, Biological,*  
1105 *and Environmental Statistics* **7**, 277–298 (2002)
- 1106 [67] King, R., Brooks, S.P.: Survival and spatial fidelity of mouflons: The  
1107 effect of location, age, and sex. *Journal of Agricultural, Biological, and*  
1108 *Environmental Statistics* **8**, 486–513 (2003)
- 1109 [68] McCrea, R.S., Morgan, B.J.T.: Multistate mark–recapture model selec-  
1110 tion using score tests. *Biometrics* **67**, 234–241 (2011) [https://arxiv.org/](https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2010.01421.x)  
1111 [abs/https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2010.](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2010.01421.x)  
1112 [01421.x](https://doi.org/10.1111/j.1541-0420.2010.01421.x). <https://doi.org/10.1111/j.1541-0420.2010.01421.x>
- 1113 [69] Burnham, K.P., Anderson, D.R.: *Model Selection and Multimodel*  
1114 *Inference*. Springer, New York (2002)
- 1115 [70] Pohle, J., Langrock, R., van Beest, F.M., Schmidt, N.M.: Selecting the  
1116 number of states in hidden Markov models: Pragmatic solutions illus-  
1117 trated using animal movement. *Journal of Agricultural, Biological, and*  
1118 *Environmental Statistics* **22**, 270–293 (2017)

- 1119 [71] Baillie, S.R., Brooks, S.P., King, R., Thomas, L.: Using a state-space  
1120 model of the British Song Thrush *Turdus philomelos* population to  
1121 diagnose the causes of a population decline. In: Thomson, D., Cooch,  
1122 E.G., Conroy, M.J. (eds.) *Modeling Demographic Processes in Marked*  
1123 *Populations*, pp. 541–561. Springer, Boston, MA (2009)
- 1124 [72] Davis, A.J., Hooten, M.B., Phillips, M.L., Doherty Jr, P.F.: An inte-  
1125 grated modeling approach to estimating gunnison sage-grouse population  
1126 dynamics: combining index and demographic data. *Ecology and Evolu-*  
1127 *tion* **4**, 4247–4257 (2014)
- 1128 [73] Bengtsson, T., Cavanaugh, J.E.: An improved akaike information crite-  
1129 rion for state-space model selection. *Computational Statistics & Data*  
1130 *Analysis* **50**, 2635–2654 (2006)
- 1131 [74] Besbeas, P., McCrea, R.S., Morgan, B.J.T.: Selecting survival age  
1132 structure for integrated population models. *Ecological modelling*, (2022)
- 1133 [75] Hooten, M.B., Hobbs, N.T.: A guide to Bayesian model selection for  
1134 ecologists. *Ecological Monographs* **85**, 3–28 (2015)
- 1135 [76] King, R., Morgan, B.J.T., Gimenez, O., Brooks, S.P.: *Bayesian Analysis*  
1136 *for Population Ecology*. CRC Press, Boca Raton (2009)
- 1137 [77] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K.,  
1138 Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., Yau,  
1139 C.: Bayesian statistics and modelling. *Nature Reviews Methods Primers*  
1140 **1**, 1–26 (2021)
- 1141 [78] Gimenez, O., Morgan, B.J.T., Brooks, S.P.: Weak identifiability  
1142 in models for mark-recapture-recovery data. In: Thomson, D.L.,

- 1143 Cooch, E.G., Conroy, M.J. (eds.) Modeling Demographic Pro-  
1144 cesses In Marked Populations, pp. 1055–1067. Springer, Boston, MA  
1145 (2009). [https://doi.org/10.1007/978-0-387-78151-8\\_48](https://doi.org/10.1007/978-0-387-78151-8_48). [https://doi.org/](https://doi.org/10.1007/978-0-387-78151-8_48)  
1146 [10.1007/978-0-387-78151-8\\_48](https://doi.org/10.1007/978-0-387-78151-8_48)
- 1147 [79] McCrea, R.S., Morgan, B.J.T., Pradel, R.: Diagnostic goodness-of-fit  
1148 tests for joint recapture and recovery models. *Journal of Agricultural,*  
1149 *Biological, and Environmental Statistics* **19**, 338–356 (2014)
- 1150 [80] McCrea, R.S., Morgan, B.J.T., Gimenez, O.: A new strategy for diag-  
1151 nostic model assessment in capture-recapture. *Journal of the Royal*  
1152 *Statistical Society. Series C (Applied Statistics)* **66**, 815–831 (2017)
- 1153 [81] Jeyam, A., McCrea, R., Pradel, R.: Assessing heterogeneity in tran-  
1154 sition propensity in multistate capture–recapture data. *Journal of*  
1155 *the Royal Statistical Society: Series C (Applied Statistics)* **69**, 413–  
1156 427 (2020) [https://arxiv.org/abs/https://rss.onlinelibrary.wiley.com/](https://arxiv.org/abs/https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12392)  
1157 [doi/pdf/10.1111/rssc.12392](https://arxiv.org/abs/https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12392). <https://doi.org/10.1111/rssc.12392>
- 1158 [82] Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R., Hooten, M.B.: A  
1159 guide to Bayesian model checking for ecologists. *Ecological Monographs*  
1160 **88**, 526–542 (2018)
- 1161 [83] Besbeas, P., Morgan, B.J.T.: Goodness-of-fit of integrated population  
1162 models using calibrated simulation. *Methods in Ecology and Evolution*  
1163 **5**, 1373–1382 (2014). <https://doi.org/10.1111/2041-210X.12279>
- 1164 [84] Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten,  
1165 M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen,  
1166 L.G., Loescher, H.W., Lunch, C.K., Pijanowski, B.C., Randerson, J.T.,



- 1167 Read, E.K., Tredennick, A.T., Vargas, R., Weathers, K.C., White, E.P.:  
1168 Iterative near-term ecological forecasting: Needs, opportunities, and  
1169 challenges. *Proceedings of the National Academy of Sciences of the*  
1170 *United States of America* **115**, 1424–1432 (2018). [https://doi.org/10.](https://doi.org/10.1073/pnas.1710231115)  
1171 [1073/pnas.1710231115](https://doi.org/10.1073/pnas.1710231115)
- 1172 [85] Buhle, E.R., Scheuerell, M.D., Cooney, T.D., Ford, M.J., Zabel, R.W.,  
1173 Thorson, J.T.: Using integrated population models to evaluate fishery  
1174 and environmental impacts on pacific salmon viability (2018). [https://](https://doi.org/10.7289/V5/TM-NWFSC-140)  
1175 [doi.org/10.7289/V5/TM-NWFSC-140](https://doi.org/10.7289/V5/TM-NWFSC-140)
- 1176 [86] The IUCN Red List: The IUCN Red List of Threatened Species. Version  
1177 2021-3. <https://www.iucnredlist.org>.
- 1178 [87] Converse, S.J., Moore, C.T., Armstrong, D.P.: Demographics of reintro-  
1179 duced populations: Estimation, modeling, and decision analysis. *Journal*  
1180 *of Wildlife Management* **77**, 1081–1093 (2013). [https://doi.org/10.1002/](https://doi.org/10.1002/jwmg.590)  
1181 [jwmg.590](https://doi.org/10.1002/jwmg.590)
- 1182 [88] Predicting harvest impact and establishment success when translocating  
1183 highly mobile and endangered species. *Journal of Applied Ecology* **00**,  
1184 1–13 (2022). <https://doi.org/10.1111/1365-2664.14219>
- 1185 [89] Neupane, N., Zipkin, E.F., Saunders, S.P., Ries, L.: Grappling with  
1186 uncertainty in ecological projections: a case study using the migratory  
1187 monarch butterfly. *Ecosphere* **13**, 03874 (2022). [https://doi.org/10.1002/](https://doi.org/10.1002/ECS2.3874)  
1188 [ECS2.3874](https://doi.org/10.1002/ECS2.3874)
- 1189 [90] DeFilippo, L.B., Buehrens, T.W., Scheuerell, M., Kendall, N.W.,  
1190 Schindler, D.E.: Improving short-term recruitment forecasts for coho

- 1191 salmon using a spatiotemporal integrated population model. *Fisheries*  
1192 *Research* **242** (2021). <https://doi.org/10.1016/j.fishres.2021.106014>
- 1193 [91] Johnson, D.S., Hoeting, J.A.: Autoregressive models for capture-  
1194 recapture data: A bayesian approach. *Biometrics* **59**, 341–350 (2003)
- 1195 [92] Codling, E.A., Plank, M.J., Benhamou, S.: Random walk models in  
1196 biology. *Journal of The Royal Society Interface* **5**, 813–834 (2008).  
1197 <https://doi.org/10.1098/RSIF.2008.0014>
- 1198 [93] Ives, A.R., Abbott, K.C., Ziebarth, N.L.: Analysis of ecological time  
1199 series with arma(p,q) models. *Ecology* **91**, 858–871 (2010). <https://doi.org/10.1890/09-0442.1>
- 1201 [94] Oppel, S., Hilton, G., Ratcliffe, N., Fenton, C., Daley, J., Gray, G., Vick-  
1202 eryl, J., Gibbons, D.: Assessing Population Viability While Accounting  
1203 for Demographic and Environmental Uncertainty
- 1204 [95] Boyce, M.S.: Population viability analysis. *Annual Review of Ecol-*  
1205 *ogy and Systematics* **23**, 481–506 (1992). [https://doi.org/10.1146/](https://doi.org/10.1146/ANNUREV.ECOLSYS.23.1.481)  
1206 [ANNUREV.ECOLSYS.23.1.481](https://doi.org/10.1146/ANNUREV.ECOLSYS.23.1.481)
- 1207 [96] Schaub, M., Kéry, M.: *Integrated Population Models: Theory and Eco-*  
1208 *logical Applications with R and JAGS*, p. 622. Academic Press, ???  
1209 (2021)
- 1210 [97] Schwacke, L., Thomas, L., Wells, R., McFee, W., Hohn, A., Mullin, K.,  
1211 Zolman, E., Quigley, B., Rowles, T., Schwacke, J.: Quantifying injury to  
1212 common bottlenose dolphins from the deepwater horizon oil spill using  
1213 an age-, sex- and class-structured population model. *Endangered Species*  
1214 *Research* **33**, 265–279 (2017). <https://doi.org/10.3354/esr00777>

- 1215 [98] Gauthier, G., Péron, G., Lebreton, J.-D., Grenier, P., van Oudenhove,  
1216 L.: Partitioning prediction uncertainty in climate-dependent population  
1217 models. *Proceedings of the Royal Society B: Biological Sciences* **283**,  
1218 20162353 (2016)
- 1219 [99] Iles, D., Jenouvrier, S.: *Projected Population Consequences of Climate*  
1220 *Change*, pp. 147–164. Oxford University Press, Oxford (2019)
- 1221 [100] Zylstra, E.R., Ries, L., Neupane, N., Saunders, S.P., Ramírez, M.I.,  
1222 Rendón-Salinas, E., Oberhauser, K.S., Farr, M.T., Zipkin, E.F.: Changes  
1223 in climate drive recent monarch butterfly dynamics. *Nature Ecology*  
1224 *and Evolution* **5**, 1441–1452 (2021). [https://doi.org/10.1038/  
1225 s41559-021-01504-1](https://doi.org/10.1038/s41559-021-01504-1)
- 1226 [101] Zylstra, E.R., Zipkin, E.F.: Accounting for sources of uncertainty when  
1227 forecasting population responses to climate change. *Journal of Animal*  
1228 *Ecology* **90**, 558–561 (2021). <https://doi.org/10.1111/1365-2656.13443>
- 1229 [102] Jenouvrier, S., Holland, M., Stroeve, J., Barbraud, C., Weimerskirch, H.,  
1230 Serreze, M., Caswell, H.: Effects of climate change on an emperor pen-  
1231 guin population: analysis of coupled demographic and climate models.  
1232 *Global change biology* **18**, 2756–2770 (2012). [https://doi.org/10.1111/J.  
1233 1365-2486.2012.02744.X](https://doi.org/10.1111/J.1365-2486.2012.02744.X)
- 1234 [103] Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee, S., Hamil,  
1235 K.A.D., Koenig, L.E., Locke, D.H., Peterson, M., Tingley, M.W.,  
1236 Wheeler, K., Youngflesh, C., Zipkin, E.F.: Working across space and  
1237 time: nonstationarity in ecological research and application. *Frontiers*  
1238 *in Ecology and the Environment* **19**, 66–72 (2021). [https://doi.org/10.  
1239 1002/fee.2298](https://doi.org/10.1002/fee.2298)

- 1240 [104] Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse,  
1241 B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays,  
1242 R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C.,  
1243 Stewart, C.V., Berger-Wolf, T.: Perspectives in machine learning for  
1244 wildlife conservation. *Nature Communications* 2022 13:1 **13**, 1–15 (2022).  
1245 <https://doi.org/10.1038/s41467-022-27980-y>
- 1246 [105] Robinson, O.J., Ruiz-Gutierrez, V., Fink, D., Meese, R.J., Holyoak, M.,  
1247 Cooch, E.G.: Using citizen science data in integrated population models  
1248 to inform conservation. *Biological Conservation* **227**, 361–368 (2018)
- 1249 [106] Johnston, A., Matechou, E., Dennis, E.B.: Outstanding challenges and  
1250 future directions for biodiversity monitoring using citizen science data.  
1251 *Methods in Ecology and Evolution* (2022). [https://doi.org/10.1111/](https://doi.org/10.1111/2041-210X.13834)  
1252 [2041-210X.13834](https://doi.org/10.1111/2041-210X.13834). In press
- 1253 [107] Fukaya, K., Kondo, N.I., ichiro S. Matsuzaki, S., Kadoya, T.: Multi-  
1254 species site occupancy modelling and study design for spatially replicated  
1255 environmental dna metabarcoding. *Methods in Ecology and Evolution*  
1256 **13**, 183–193 (2022). <https://doi.org/10.1111/2041-210X.13732>
- 1257 [108] Hodgson, J.C., Baylis, S.M., Mott, R., Herrod, A., Clarke, R.H.: Pre-  
1258 cision wildlife monitoring using unmanned aerial vehicles. *Scientific*  
1259 *Reports* 2016 6:1 **6**, 1–7 (2016). <https://doi.org/10.1038/srep22574>
- 1260 [109] Estimating animal population density using passive acoustics. *Biological*  
1261 *Reviews* **88**, 287–309 (2013). <https://doi.org/10.1111/BRV.12001>
- 1262 [110] Duporge, I., Isupova, O., Reece, S., Macdonald, D.W., Wang, T.: Using  
1263 very-high-resolution satellite imagery and deep learning to detect and

- 1264 count african elephants in heterogeneous landscapes. *Remote Sensing in*  
1265 *Ecology and Conservation* **7**, 369–381 (2021). [https://doi.org/10.1002/](https://doi.org/10.1002/RSE2.195)  
1266 [RSE2.195](https://doi.org/10.1002/RSE2.195)
- 1267 [111] Ross, B.E., Weegman, M.D.: Relative effects of sample size, detection  
1268 probability, and study duration on estimation in integrated population  
1269 models. *Ecological Applications*, 2686 (2022). [https://doi.org/10.1002/](https://doi.org/10.1002/EAP.2686)  
1270 [EAP.2686](https://doi.org/10.1002/EAP.2686)
- 1271 [112] Zhao, Q.: On the sampling design of spatially explicit integrated popu-  
1272 lation models. *Methods in Ecology and Evolution* **11**, 1207–1220 (2020).  
1273 <https://doi.org/10.1111/2041-210X.13457>
- 1274 [113] Borysiewicz, R.S., Morgan, B.J.T., Hénau, V., Bregnballe, T., Lebre-  
1275 ton, J.-D., Gimenez, O.: An integrated analysis of multisite recruit-  
1276 ment, mark-recapture-recovery and multisite census data. In: Thomson,  
1277 D.L., Cooch, E.G., Conroy, M.J. (eds.) *Modeling Demographic Pro-*  
1278 *cesses In Marked Populations*, pp. 579–591. Springer, Boston, MA  
1279 (2009). [https://doi.org/10.1007/978-0-387-78151-8\\_25](https://doi.org/10.1007/978-0-387-78151-8_25). [https://doi.org/](https://doi.org/10.1007/978-0-387-78151-8_25)  
1280 [10.1007/978-0-387-78151-8\\_25](https://doi.org/10.1007/978-0-387-78151-8_25)
- 1281 [114] Barraquand, F., Gimenez, O.: Fitting stochastic predator–prey models  
1282 using both population density and kill rate data. *Theoretical Population*  
1283 *Biology* **138**, 1–27 (2021)
- 1284 [115] Catchpole, E.A., Morgan, B.J.T.: Detecting parameter redun-  
1285 dancy. *Biometrika* **84**, 187–196 (1997) [https://arxiv.org/abs/https://](https://arxiv.org/abs/https://academic.oup.com/biomet/article-pdf/84/1/187/669241/84-1-187.pdf)  
1286 [academic.oup.com/biomet/article-pdf/84/1/187/669241/84-1-187.pdf](https://arxiv.org/abs/https://academic.oup.com/biomet/article-pdf/84/1/187/669241/84-1-187.pdf).  
1287 <https://doi.org/10.1093/biomet/84.1.187>

- 1288 [116] Choquet, R., Cole, D.J.: A hybrid symbolic-numerical method for deter-  
1289 mining model structure. *Mathematical Biosciences* **236**, 117–125 (2012).  
1290 <https://doi.org/10.1016/j.mbs.2012.02.002>
- 1291 [117] Besbeas, P., Lebreton, J.-D., Morgan, B.J.T.: The efficient integration  
1292 of abundance and demographic data. *Journal of the Royal Statistical*  
1293 *Society. Series C (Applied Statistics)* **52**, 95–102 (2003)
- 1294 [118] King, R., Sarzo, B., Elvira, V.: Large data and (not even very) complex  
1295 ecological models: When worlds collide. arXiv.2205.07261 (2022)
- 1296 [119] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betan-  
1297 court, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: A probabilistic  
1298 programming language. *Journal of Statistical Software* **76** (2017)
- 1299 [120] de Valpine, P., Turek, D., Paciorek, C.J., Anderson-Bergman, C.,  
1300 Lang, D.T., Bodik, R.: Programming with models: Writing statisti-  
1301 cal algorithms for general model structures with nimble. *Journal of*  
1302 *Computational and Graphical Statistics* **26**, 403–413 (2017)
- 1303 [121] Ponisio, L.C., de Valpine, P., Michaud, N., Turek, D.: One size does  
1304 not fit all: Customizing MCMC methods for hierarchical models using  
1305 NIMBLE. *Ecology and Evolution* **10**, 2385–2416 (2020)
- 1306 [122] Turek, D., Milleret, C., Ergon, T., Brøseth, H., Dupont, P., Bischof,  
1307 R., de Valpine, P.: Efficient estimation of large-scale spatial cap-  
1308 ture–recapture models. *Ecosphere* **12**, 03385 (2021)
- 1309 [123] ISO: ISO International Standard ISO/IEC 14882:2020(E) – Program-  
1310 ming Language C++

- 1311 [124] Sisson, S.A., Fan, Y., Beaumont, M.: Handbook of Approximate  
1312 Bayesian Computation. CRC Press, Boca Raton (2018)
- 1313 [125] Wood, S.N.: Statistical inference for noisy nonlinear ecological dynamic  
1314 systems. *Nature* **466**, 1102–1104 (2010)
- 1315 [126] McElreath, R.: Statistical Rethinking: a Bayesian Course with Examples  
1316 in R And Stan. Chapman & Hall/CRC texts in statistical science series.  
1317 CRC Press/Taylor & Francis Group, Boca Raton (2016)
- 1318 [127] Plard, F., Turek, D., Schaub, M.: Consequences of violating assumptions  
1319 of integrated population models on parameter estimates. *Environmental  
1320 and Ecological Statistics* **28**, 667–695 (2021)
- 1321 [128] Jeyam, A., McCrea, R.S., Pradel, R.: A test for the underlying state-  
1322 structure of hidden Markov models: Partially observed capture-recapture  
1323 data. *Frontiers in Ecology and Evolution* **9** (2021). [https://doi.org/10.  
1324 3389/fevo.2021.598325](https://doi.org/10.3389/fevo.2021.598325)
- 1325 [129] Durbin, J., Koopman, S.J.: Time Series Analysis by State Space  
1326 Methods. Oxford University Press, Oxford (2001)
- 1327 [130] Dietze, M.C.: Prediction in ecology: a first-principles framework. *Ecolog-  
1328 ical Applications* **27**, 2048–2060 (2017). [https://doi.org/10.1002/EAP.  
1329 1589](https://doi.org/10.1002/EAP.1589)
- 1330 [131] Nichols, J.D., Runge, M.C., Johnson, F.A., Williams, B.K.: Adaptive  
1331 harvest management of north american waterfowl populations: A brief  
1332 history and future prospects. *Journal of Ornithology* **148** (2007). [https:  
1333 //doi.org/10.1007/s10336-007-0256-8](https://doi.org/10.1007/s10336-007-0256-8)

- 1334 [132] Zurell, D., Graham, C.H., Gallien, L., Thuiller, W., Zimmermann, N.E.:  
1335 Long-distance migratory birds threatened by multiple independent risks  
1336 from global change. *Nature Climate Change* **8**, 992–996 (2018). <https://doi.org/10.1038/s41558-018-0312-9>  
1337
- 1338 [133] Wainwright, T.C.: Ephemeral relationships in salmon forecasting: A cau-  
1339 tionary tale. *Progress in Oceanography* **193** (2021). [https://doi.org/10.](https://doi.org/10.1016/j.pocean.2021.102522)  
1340 [1016/j.pocean.2021.102522](https://doi.org/10.1016/j.pocean.2021.102522)
- 1341 [134] Lieury, N., Devillard, S., Besnard, A., Gimenez, O., Hameau, O., Pon-  
1342 chon, C., Millon, A.: Designing cost-effective capture-recapture surveys  
1343 for improving the monitoring of survival in bird populations. *Biologi-*  
1344 *cal Conservation* **214**, 233–241 (2017). [https://doi.org/10.1016/j.biocon.](https://doi.org/10.1016/j.biocon.2017.08.011)  
1345 [2017.08.011](https://doi.org/10.1016/j.biocon.2017.08.011)
- 1346 [135]
- 1347 [136] Baker, C.M., Gordon, A., Bode, M.: Ensemble ecosystem modeling for  
1348 predicting ecosystem response to predator reintroduction. *Conservation*  
1349 *Biology* **31**, 376–384 (2017). <https://doi.org/10.1111/cobi.12798>
- 1350 [137] Baker, C.M., Bode, M., Dexter, N., Lindenmayer, D.B., Foster, C.,  
1351 MacGregor, C., Plein, M., McDonald-Madden, E.: A novel approach  
1352 to assessing the ecosystem-wide impacts of reintroductions. *Ecological*  
1353 *Applications* **29** (2019). <https://doi.org/10.1002/eap.1811>
- 1354 [138] Spence, M.A., Blanchard, J.L., Rossberg, A.G., Heath, M.R., Heymans,  
1355 J.J., Mackinson, S., Serpetti, N., Speirs, D.C., Thorpe, R.B., Blackwell,  
1356 P.G.: A general framework for combining ecosystem models. *Fish and*  
1357 *Fisheries* **19**, 1031–1042 (2018). <https://doi.org/10.1111/FAF.12310>



- 1358 [139] Plard, F., Turek, D., Gruebler, M.U., Schaub, M.: Ipm2: toward bet-  
1359 ter understanding and forecasting of population dynamics. *Ecological*  
1360 *Monographs* **89**, 01364 (2019). <https://doi.org/10.1002/ECM.1364>

Challenge	Existing Work	Future Avenues
Model specification	Ecological knowledge (system process); Sampling methods (observation process); Dependence structure; Parameter interpretation; Confounded parameters.	New forms of data; Accounting for variability in quality/quantity of different datasets; Dealing with data across different geographical/temporal scales; Model identifiability and parameter redundancy; Multi-species and associated issues.
Computational	Optimisation of computational times for each component separately through likelihood marginalisation and approximation; Formulation of IPMs as HMMs and use of associated machinery.	Approximate likelihood approaches; Scalability of algorithms; General and easy-to-use computational solutions; Alternative software solutions for IPM implementation (besides Jags).
Model assessment	Step-wise model selection; most informative data set; posterior model probabilities; RJMCMC; diagnostic goodness-of-fit tests; Bayesian p-values; calibrated simulation.	Regularisation methods; HMM methodology; forecast variance; approximate likelihood approaches; software development.
Forecasting	Time-series methods; Population viability analysis; Climate projections; Ensemble models.	Quantifying and reducing uncertainty; Optimising sampling design; Combining IPM with ensemble models; Coupling IPMs with integral projection models (IPM <sup>2</sup> ).

**Table 1** Table summarising four practical challenges associated with the application of IPMs to ecological data, the existing work which addresses them and the future avenues of research.