Zero-inflated count distributions for capture-mark-reencounter data

Thomas V. Riecke[1,2], Dan Gibson[3], James S. Sedinger[4], and Michael Schaub[1]

[1]Swiss Ornithological Institute, CH-6204 Sempach, Switzerland

[2]thomasvanceriecke@gmail.com

[3]Warner College of Natural Resources, Colorado State University, Fort Collins, Colorado, USA

[4]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada, USA

April 15, 2022

1 **Abstract**

2 The estimation of demographic parameters is a key component of evolutionary demography and

3 conservation biology. Capture-mark-recapture methods have served as a fundamental tool for

4 estimating demographic parameters. The accurate estimation of demographic parameters in

5 capture-mark-recapture studies depends on accurate modeling of the observation process. Classic

6 capture-mark-recapture models typically model the observation process as a Bernoulli or

7 categorical trial with some detection probability conditional on a marked individual's availability

8 for detection (e.g., alive, or alive and present in a study area). Alternatives to this approach are

9 underused, but may have great utility in capture-recapture studies. In this paper we explore a

10 simple concept: *in the same way that counts contain more information about abundance than*

11 *simple detection/non-detection data, the number of encounters of individuals during observation*

12 *occasions contain more information about the observation process than detection/non-detection*

13 *data for individuals during the same occasion*. Rather than using Bernoulli or categorical

14 distributions to estimate detection probability, we demonstrate the application of zero-inflated

15 Poisson and gamma-Poisson distributions. This allows for inference on availability for encounter

16 (i.e., temporary emigration), as well as a wide variety of parameterizations for heterogeneity in

17 the observation process. We demonstrate that this approach can accurately recover demographic

18 and observation parameters in the presence of individual heterogeneity in detection probability,

19 and discuss some potential future extensions of this method.

20 **Key words:** Bayesian, capture-mark-recapture, gamma-Poisson, individual heterogeneity,

21 mark-resight, robust design, temporary emigration, zero-inflation

Introduction

The estimation of demographic parameters is fundamental to successful conservation and

evolutionary ecology. Since their initial development, capture-mark-recapture (hereafter, CMR)

models have been used to estimate demographic parameters such as apparent survival (Cormack,

1964; Jolly, 1965; Seber, 1965), true survival and site fidelity (Burnham, 1993), transitions

among discrete strata (Brownie et al., 1993), temporary emigration or breeding probability

(Kendall et al., 1995, 1997), recruitment (Pradel, 1996), and the spatial distribution of organisms

(Royle and Young, 2008; Royle et al., 2013). Parameter estimates from CMR models are often

used as vital components of population models (Caswell, 2000; Schaub and Kéry, 2021), and for

examining life-history trade-offs (Cam et al., 2002). CMR models typically consist of two

primary components: 1) a model of latent biological processes (i.e., survival, movement among

populations, emigration, disease dynamics), and 2) a model of the observation of uniquely

identifiable individuals. Models of both latent biological and observation processes typically take

the form of categorical or Bernoulli distributions, and individuals are grouped into discrete groups

or states (e.g., alive or dead, observed or not observed).

Heterogeneity among 'uniquely identifiable' (hereafter, marked) organisms in both

biological processes (e.g., Cam et al. 2002; Pledger and Schwarz 2002) and observation

probability (e.g., Pollock 1982; Pledger 2005) has long been recognized as a central challenge in

CMR modeling (Otis et al., 1978). In a seminal paper, Pollock (1982) proposed that heterogeneity

in detection might be accounted for by sub-dividing primary occasions into multiple secondary

occasions. Similarly, Fletcher (1994) developed a method for modeling the probability of

encounter of individuals as a function of the number of unique resights of that individual during

3

44 the previous occasion. Shortly thereafter, Kendall and others (Kendall et al., 1995, 1997)

45 expanded the method developed by Pollock (1982) to estimate availability for encounter (i.e.,

46 zero-inflation) by partitioning primary occasions into shorter secondary occasions, assuming

47 closure among secondary occasions within a primary occasion, and estimating probabilities of

48 temporary emigration from the study area. Since that time, methods have been developed to

49 estimate individual detection probabilities using random effects (Clark et al., 2005; Royle and

50 Dorazio, 2008) or mixtures (Pledger, 2000; Pledger et al., 2003). More recent efforts have

51 simultaneously used information about marked organism location and the locations of sampling

52 efforts to model spatial variation in reencounter probability (Royle and Young, 2008; Royle et al.,

53 2013). However, the estimation of heterogeneity in the observation process remains a key

54 challenge in CMR studies, and the continued development of alternative approaches is critical for

55 parameter estimation.

56   Heterogeneity in detection of marked organisms is often driven by two primary processes.

57 The first is whether or not an individual is even present within the bounds of the study area (i.e.,

58 temporary emigration or zero-inflation; Kendall et al. 1995; Schaub et al. 2004). The second is

59 variation among the latent encounter probabilities of individuals that are present. This latent

60 heterogeneity can be affected by factors such as variation in individual behavior, life stage, and

61 location relative to sampling effort (Royle and Young, 2008). When primary occasions extend

62 over multiple days, weeks, or months, this can lead to some individuals being encountered many

63 times while others are rarely, if ever, detected. The key concept in this paper is that *in the same*

64 *way that counts contain more information about the abundance of a population than simple*

65 *detection/non-detection data, the number of encounters of marked individuals may contain more*

66 *information about the observation process than detection/non-detection data* (e.g., McClintock

4

67 and White 2009; McClintock et al. 2009, 2019). Thus, rather than summarizing

68 capture-reencounter data using ones (encountered) and zeroes (not encountered) during a primary

69 occasion or multiple secondary occasions, capture-reencounter data can also be summarized as

70 counts of the number of times each marked individual was encountered during a primary occasion

71 (McClintock and White, 2009; McClintock et al., 2019). The number of encounters can then be

72 modelled using a variety of discrete distributions, such as the Poisson and negative binomial

73 distributions. If model assumptions are met, this approach provides a flexible and useful

74 extension to existing CMR models, and may improve upon existing tools to estimate

75 heterogeneity in encounter probability among individuals. Notably, improved estimates of

76 heterogeneity in the observation process lead to improved estimates of demographic parameters.

77 In this paper, we 1) demonstrate the use of this approach with simulated data, 2) describe

78 potential benefits relative to more traditional approaches, 3) demonstrate several approaches for

79 modeling individual heterogeneity in encounter probability, and 4) discuss possible future

80 extensions and uses of this parameterization.

81 Methods

82 We simulated 250 CMR datasets, each with ten primary occasions ($T = 10$). For each simulation,

83 we released 25 marked individuals in the first through ninth primary occasions, for a total of 225

84 released individuals ($I = 225$). We simulated the latent state of each individual ($z_{i,t}$; 1: alive, 0:

85 dead) from occasion to occasion as, $z_{i,t} \sim \text{Bernoulli}(z_{i,t-1}\phi)$, given a simulated survival

86 probability, $\phi \sim \text{beta}(40, 10)$. If an individual was alive in occasion t, we simulated its

87 availability for encounter ($a_{i,t}$; 1: available, 0: unavailable) given simulated Markovian (Kendall

88 et al., 1997) probabilities of availability for encounter ($\boldsymbol{\gamma}$),

$$a_{i,t} \sim \text{Bernoulli}(z_{i,t} \times \gamma_1 \times (1 - a_{i,t\text{-}1}) + z_{i,t} \times \gamma_2 \times a_{i,t\text{-}1}),$$

$$\gamma_1 \sim \text{beta}(10, 20), \tag{1}$$

$$\gamma_2 \sim \text{beta}(20, 10).$$

89 These probabilities are similar to temporary emigration parameters described by Kendall et al.

90 (1997), such that $\gamma_2 = 1 - \gamma''$ and $\gamma_1 = 1 - \gamma'$. During each primary occasion, we sampled

91 individuals that were available for detection for 21 consecutive days ($J = 21$) given simulated

92 individual random variation in daily detection probability (d_i; eq. 2). Thus, the simulated

93 capture-recapture data form a 3-dimensional array (Y) with dimensions $I \times T \times J$,

$$y_{i,t,j} \sim \text{Bernoulli}(a_{i,t} \times d_i),$$

$$d_i \sim \text{beta}(\mu_\delta \times \left(\frac{1}{\sigma_\delta^2}\right), (1 - \mu_\delta) \times \left(\frac{1}{\sigma_\delta^2}\right)),$$

$$\mu_\delta \sim \text{beta}(10, 90), \tag{2}$$

$$\sigma_\delta \sim \text{gamma}(5, 50).$$

94 We then summarized the daily CMR data for analysis with four different model types: 1) a

95 Cormack-Jolly-Seber model where the secondary captures are ignored (CJS; Cormack 1964; Jolly

96 1965; Seber 1965), 2) a robust design model (RD; Kendall et al. 1995, 1997), and two

97 capture-recapture models with count-based observation likelihoods, 3) a zero-inflated Poisson

98 (ZIP), and 4) a zero-inflated gamma-Poisson with heterogeneity in the number of encounters per

99 individual (ZIGP). To summarize the CMR data (M) for a CJS model, we constructed an $I \times T$

100 matrix and filled the matrix as a function of whether or not an individual was observed on any day

101 during a primary occasion,

$$m_{i,t} \sim \begin{cases} 1, & \text{if } \sum_{j=1}^{21} y_{i,t,j} \geq 1 \\ \\ 0, & \text{otherwise} \end{cases} . \tag{3}$$

102 To summarize the robust design encounter data (R) for the robust design capture-reencounter

103 model, we subdivided each 21-day long primary occasion into three 7-day long secondary

104 occasions ($K = 3$). If an individual was observed on any day of a week in a secondary occasion,

105 then that secondary occasion ($r_{i,t,k}$) equaled one. If an individual was not observed on any day

106 during a specific secondary occasion, then $r_{i,t,k} = 0$. Finally, we summarized the counts of

107 reencounters by individual and primary occasion by simply summing the total number of

108 encounters of each individual during each primary occasion, $c_{i,t} = \sum_{j=1}^{21} y_{i,t,j}$.

109   In the same way that the data were generated, all four capture-recapture models share a

110 common likelihood for the survival process. The latent state of each individual during each

111 occasion ($z_{i,t}$) was modeled as a function of the individual's latent state in the previous occasion

112 ($z_{i,t-1}$) and a survival probability (ϕ), $z_{i,t} \sim$ Bernoulli($z_{i,t-1} \times \phi$). A vague prior was used for

113 survival, $\phi \sim$ beta($1, 1$). For the CJS model, we then simply modeled the primary occasions

114 encounter data (M) as a function of the individual's latent state and a detection probability (p),

115 $m_{i,t} \sim$ Bernoulli($z_{i,t} \times p$). We specified a vague prior for detection probability $p \sim$ Beta($1, 1$).

116 For the remaining three models, we also estimated whether an individual was available for

117 detection ($a_{i,t}$) given its previous state ($a_{i,t-1}$) and vague priors for Markovian probabilities of

118 availability for encounter (γ; Kendall et al. 1997).

$$a_{i,t} \sim \text{Bernoulli}(z_{i,t} \times \gamma_1 \times (1 - a_{i,t-1}) + z_{i,t} \times \gamma_2 \times a_{i,t-1}),$$

$$\boldsymbol{\gamma} \sim \text{beta}(1, 1). \tag{4}$$

119 For the robust design model, we modeled whether or not each individual was detected during

120 each secondary occasion as a function of its latent availability for detection during the primary

121 occasion ($a_{i,t}$) and a secondary occasion detection probability (p). We then derived primary

122 occasion detection probability (p^*) from the secondary occasion detection probabilities for

123 comparison of parameter estimates among models,

$$r_{i,t,j} \sim \text{Bernoulli}(a_{i,t} \times p),$$

$$p \sim \text{Beta}(1, 1),$$

$$p^* = 1 - (1 - p)^3. \tag{5}$$

124 For the zero-inflated Poisson model, we model the total number of encounters of each individual

125 during each primary occasion ($c_{i,t}$) given availability for detection ($a_{i,t}$) an expected mean number

126 of encounters per individual per primary occasion (ϵ),

$$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon),$$

$$\epsilon \sim \text{Gamma}(1, 1). \tag{6}$$

127 For the zero-inflated Gamma-Poisson model with heterogeneity in the number of expected

128 observations per individual, we modeled the number of encounters of each individual during each

129 primary occasion ($c_{i,t}$) given availability for detection ($a_{i,t}$), the mean expected number of

encounters per individual (ϵ), and an overdispersion parameter (θ),

$$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon \times h_i),$$

$$\epsilon \sim \text{gamma}(1, 1),$$

$$h_i \sim \text{gamma}(\theta, \theta), \tag{7}$$

$$\theta \sim \text{uniform}(0, 250).$$

This parameterization is similar to Gamma-Poisson formulations of the negative binomial

distribution (Greene, 2008), however, here we assume heterogeneity among individuals, not

observations (Table 1). We called JAGS (Plummer, 2003) from R (R Core Team, 2018) using the

`jagsUI` package (Kellner, 2016). We sampled three MCMC chains of 50,000 iterations with an

adaptive phase of 1,000 iterations. We discarded the first 10,000 iterations and retained every

tenth saved iteration. We assessed convergence visually, and chains converged acceptably.

Results

Estimates of survival (ϕ) were biased for CJS models (MSD = -0.047; Coverage = 0.464), but

constant and calibrated for RD (MSD = -0.003; Coverage = 0.940), ZIP (MSD = -0.002;

Coverage = 0.948), and ZIGP (MSD = 0.001; Coverage = 0.948) CMR models (Figure 1; Table

2). Estimates of availability for encounter given previous availability for encounter ($\gamma_2|a_{i,t-1} = 1$)

were slightly underestimated by RD (MSD = -0.020; Coverage = 0.892) and ZIP (MSD = -0.013;

Coverage = 0.896) models, but near truth for the ZIGP (MSD = 0.006; Coverage = 0.936) CMR

model (Figure 2; Table 2). Estimates of availability for encounter given previous unavailability

for encounter ($\gamma_1|a_{i,t-1} = 0$) were slightly overestimated by RD (MSD = 0.018; Coverage =

0.956), ZIP (MSD = 0.015; Coverage = 0.964), and ZIGP (MSD = 0.019; Coverage = 0.976)

147 CMR models, but coverage was adequate. Estimates of detection probability (p) or the average

148 number of reencounters per individual (ϵ) exhibited poor coverage (Figure 3; Table 2) for the RD

149 (MSD = 0.009; Coverage = 0.832) CMR model, were overestimated with poor coverage with the

150 ZIP (MSD = 0.078; Coverage = 0.764) CMR model, and near truth with the ZIGP (MSD = 0.002;

151 Coverage = 0.928) CMR model. The simulated individual heterogeneity in encounter probability

152 (σ_δ) in the data was positively correlated with dispersion in the count data (C; Figure 4). The

153 overdispersion parameter (θ) in the ZIGP model accounted for some of this overdispersion

154 (Figure 4), improving coverage and constancy for ZIGP models relative to other model types. ZIP

155 and ZIGP models were computationally less expensive than RD models (Figure 4).

Discussion

157 We demonstrate that CMR models parameterized with zero-inflated count distributions can

158 function much like robust design CMR models. Estimates of survival probability from RD, ZIP,

159 and ZIGP models were centered around truth, while estimates of survival from the CJS model

160 were consistently low relative to truth. Further, the use of these model types may simplify the

161 parameterization of models that account for heterogeneity in encounter probability among

162 individuals, and improve computational efficiency (Figure 4). We see substantial utility for these

163 parameterizations in a variety of scenarios, particularly when resight and recapture efforts

164 co-occur, or occur intermittently (i.e., periods of captures are followed by periods of resights). For

165 instance, non-breeding resights of individuals at wintering or stopover sites may provide an

166 excellent system to model the total number of encounters rather than simple

167 detection/non-detection data.

168   As we demonstrate herein, this approach may be particularly useful when unobservable

169 states exist, as counts of reencounters allow for the estimation of a zero-inflation parameter (i.e.,

170 availability for detection), which may be biologically analogous to breeding probability or

171 presence at a stopover or wintering site. Count parameterizations might also be used as part of a

172 robust design model; one or more secondary occasions may be estimated from some count

173 distribution, and others from a more typical Bernoulli distribution. The inherent flexibility of

174 programs like JAGS (Plummer, 2003) and NIMBLE (de Valpine et al., 2017), and ample literature

175 on capture-reencounter parameterizations should lead to a wide array of extensions of these

176 model types, and their incorporation into joint models, such as integrated population models

177 (Schaub and Kéry, 2021).

178   Critically, the use of these model types also has advantages for estimating heterogeneity in

179 detection probability among individuals. Estimating heterogeneity in probabilities from a small

180 number of Bernoulli trials can be challenging (Fay et al., 2021). Summarizing mark-reencounter

181 data as counts of encounters may provide additional information for estimating latent

182 heterogeneity among individuals or estimating mixtures (e.g., Pledger et al. 2003). For example,

183 rather than the heterogeneity parameterization explored in this paper, one might specify a mixture

184 distribution for the number of encounters per individual. Individual covariates can be

185 incorporated simply by modeling the expected number of encounters with a log-link function. We

186 anticipate a variety of other parameterizations might be useful as well (Table 1), and that

187 simulation work may reveal more effective parameterizations than those described herein. For

188 instance, recent research has demonstrated that a similar approach can be effective for addressing

189 'false-positives' in re-encounter data (Rakhimberdiev et al., 2022).

190   As with the use of any model, violations of model assumptions will lead to inaccurate

191 parameter estimates. Much like the robust design (Kendall et al., 1995, 1997), we caution against

192 the use of these models when encounters are conditional on previous encounters within a season.

193 As a particularly problematic example, if the nest of a marked animal is discovered and the

194 animal is then observed repeatedly while visiting the nest, this would serve as an additional type

195 of zero-inflation (i.e., nesting in the study area is a Bernoulli trial, the discovery of the nest is a

196 Bernoulli trial, and the subsequent visits are a product of study design and nest monitoring

197 protocols, not a random encounter process). We expect that other types of heterogeneity are

198 common in CMR data. For example, the number of encounters might be right truncated if

199 observers cease recording reencounters of individuals that have already been encountered

200 multiple times. Thus, we strongly encourage careful thought about how monitoring protocols

201 might affect the distribution of encounters of each individual when applying these models to data,

202 and discourage using this approach without explicit information about monitoring protocols.

203   The use of the Poisson distribution requires the assumption that the mean and the variance

204 are equal. When the encounter data are under- or over-dispersed, this can lead to respective under-

205 or over-estimation of the expected number of encounters per individual. Similarly, the probability

206 of availability for encounter will be over- or under-estimated given under- or over-dispersion of

207 the encounter data (Figure 4). While over-dispersion can be modeled simply using

208 gamma-Poisson mixture (demonstrated herein) or negative binomial distributions (Table 1),

209 under-dispersion requires the use of more complex distributions such as the

210 Conway-Maxwell-Poisson (Conway and Maxwell, 1962; Lynch et al., 2014). We suggest that

211 additional simulation work is required to fully understand the benefits and costs associated with

212 using alternative distributions.

213   While we have demonstrated in this paper that count-based observation parameterizations

214 can be useful for capture-mark-reencounter studies, much remains to be learned. For example,

215 careful thought will be required for developing appropriate priors (e.g., Northrup and Gerber

216 2018), and empirical research may reveal unforeseen problems. Future simulation work might

217 assess the impacts of priors on inference, further examine the impacts of over- and

218 under-dispersion, and explore various other capture-recapture parameterizations and count

219 distributions.

Acknowledgements

221 We thank David N. Koons and Madeleine G. Lohman for helpful discussion of model

222 parameterizations.

Author Contributions

224 TVR, DG, JSS, and MS conceived the central ideas behind the manuscript. TVR led the writing

225 of the manuscript and the simulation study, with important contributions from DG, JSS, and MS.

226 All authors contributed critically to manuscript editing.

Data Accessibility

228 The R script for simulating and analyzing the capture-reencounter data will be archived at the

229 Swiss Ornithological Institute's Open Repository and Archive (vogelwarte.ch).

References

Brownie, C., J. E. Hines, J. D. Nichols, K. H. Pollock, and J. B. Hestbeck. 1993. Capture-recapture studies for multiple strata including non-markovian transitions. Biometrics, pages 1173–1187.

Burnham, K. 1993. A theory for combined analysis of ring recovery and recapture data. In J. D. Lebreton and P. M. North, editors, Marked individuals in the study of bird populations, pages 199–213. Springer, Birkhauser, Basel, Switzerland.

Cam, E., W. A. Link, E. G. Cooch, J.-Y. Monnat, and E. Danchin. 2002. Individual covariation in life-history traits: seeing the trees despite the forest. The American Naturalist, **159**:96–105.

Caswell, H. 2000. Matrix population models. Sinauer, Sunderland, MA, USA.

Clark, J. S., G. Ferraz, N. Oguge, H. Hays, and J. DiCostanzo. 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology, **86**:2232–2244.

Conway, R. and W. Maxwell. 1962. A queuing model with state dependent service rates. Journal of Industrial Engineering, **12**:132–136.

Cormack, R. 1964. Estimates of survival from the sighting of marked animals. Biometrika, **51**:429–438.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. Programming with models: writing statistical algorithms for general model structures with nimble. Journal of Computational and Graphical Statistics, **26**:403–413.

Fay, R., M. Authier, S. Hamel, S. Jenouvrier, M. Van De Pol, E. Cam, J.-m. Gaillard, N. G. Yoccoz, P. Acker, A. Allen, et al. 2021. Quantifying fixed individual heterogeneity in

252 demographic parameters: Performance of correlated random effects for bernoulli variables.

253 Methods in Ecology and Evolution.

254 Fletcher, D. J. 1994. A mark-recapture model in which sighting probability depends on the

255 number of sightings on the previous occasion. In D. J. Fletcher and B. F. Manly, editors,

256 Statistics in ecology and environmental monitoring, pages 105–110. University of Otago Press,

257 Dunedin, NZ.

258 Greene, W. 2008. Functional forms for the negative binomial model for count data. Economics

259 Letters, **99**:585–590.

260 Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and

261 immigration-stochastic model. Biometrika, **52**:225–247.

262 Kellner, K. 2016. jagsUI: A wrapper Around 'rjags' to streamline 'JAGS' analyses. Version 1.5.0.

263 Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using

264 capture–recapture data with Pollock's robust design. Ecology, **78**:563–578.

265 Kendall, W. L., K. H. Pollock, and C. Brownie. 1995. A likelihood-based approach to

266 capture-recapture estimation of demographic parameters under the robust design. Biometrics,

267 **51**:293–308.

268 Kéry, M. and J. A. Royle. 2015. Applied Hierarchical Modeling in Ecology: Analysis of

269 distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static

270 Models. Academic Press.

271 Lynch, H. J., J. T. Thorson, and A. O. Shelton. 2014. Dealing with under-and over-dispersed

272 count data in life history, spatial, and community ecology. Ecology, **95**:3173–3180.

273 McClintock, B. T. and G. C. White. 2009. A less field-intensive robust design for estimating

274 demographic parameters with mark–resight data. Ecology, **90**:313–320.

McClintock, B. T., G. C. White, M. F. Antolin, and D. W. Tripp. 2009. Estimating abundance using mark–resight when sampling is with replacement or the number of marked individuals is unknown. Biometrics, **65**:237–246.

McClintock, B. T., G. C. White, and M. A. Pryde. 2019. Improved methods for estimating abundance and related demographic parameters from mark-resight data. Biometrics, **75**:799–809.

Northrup, J. M. and B. D. Gerber. 2018. A comment on priors for Bayesian occupancy models. PLOS ONE, **13**:e0192819.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs, pages 3–135.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics, **56**:434–442.

Pledger, S. 2005. The performance of mixture models in heterogeneous closed population capture–recapture. Biometrics, **61**:868–873.

Pledger, S., K. H. Pollock, and J. L. Norris. 2003. Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. Biometrics, **59**:786–794.

Pledger, S. and C. J. Schwarz. 2002. Modelling heterogeneity of survival in band-recovery data using mixtures. Journal of Applied Statistics, **29**:315–327.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing, volume 124. Vienna, Austria.

Pollock, K. H. 1982. A capture-recapture design robust to unequal probability of capture. The Journal of Wildlife Management, **46**:752–757.

298 Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and

299   population growth rate. Biometrics, **52**:703–709.

300 R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation

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

302 Rakhimberdiev, E., J. Karagicheva, A. Saveliev, A. J. Loonstra, M. A. Verhoeven, J. C.

303   Hooijmeijer, M. Schaub, and T. Piersma. 2022. Misidentifacation errors in reencounters result

304   in biased estimates of survival probability from CJS models: evidence and a solution using the

305   robust design. Methods in Ecology and Evolution, **13**.

306 Royle, J. A., R. B. Chandler, R. Sollmann, and B. Gardner. 2013. Spatial capture-recapture.

307   Academic Press, Waltham, MA.

308 Royle, J. A. and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the

309   analysis of data from populations, metapopulations and communities. Elsevier.

310 Royle, J. A. and K. V. Young. 2008. A hierarchical model for spatial capture–recapture data.

311   Ecology, **89**:2281–2289.

312 Schaub, M., O. Gimenez, B. R. Schmidt, and R. Pradel. 2004. Estimating survival and temporary

313   emigration in the multistate capture–recapture framework. Ecology, **85**:2107–2113.

314 Schaub, M. and M. Kéry. 2021. Integrated population models: Theory and ecological

315   applications with R and JAGS. Academic Press, Waltham, MA.

316 Seber, G. A. 1965. A note on the multiple-recapture census. Biometrika, **52**:249–259.

317 # Tables

Table 1. Potential parameterizations for zero-inflated count distribution based capture-reencounter models, where $c_{i,t}$ is the number of encounters of individual i during occasion t, $a_{i,t}$ is an individual's availability for encounter ($a_{i,t} = 1$ indicates available; $a_{i,t} = 0$ indicates unavailable), and ϵ is the number of expected encounters of an individual. We explicitly test parameterizations 1 and 2 in this paper, but please note that a truly vast variety of potential parameterizations exists, and see Greene (2008), Lynch et al. (2014), Kéry and Royle (2015), and McClintock et al. (2009, 2019) for further reading.

Parameterization	Model and Priors
1. Poisson	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon)$
	$\epsilon \sim \text{gamma}(1,1)$
2. Gamma-Poisson with individual heterogeneity	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon \times h_i)$
	$\epsilon \sim \text{gamma}(1,1)$
	$h_i \sim \text{gamma}(\theta, \theta)$
	$\theta \sim \text{uniform}(0, 250)$
3. Poisson with 3 categorical mixtures	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon_{\pi_i})$
	$\epsilon_j \sim \text{gamma}(1,1)$
	$\pi_i \sim \text{categorical}(\boldsymbol{\theta})$
	$\boldsymbol{\theta} \sim \text{Dirichlet}(1,1,1)$
4. Alternative Gamma-Poisson with individual heterogeneity	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times h_i)$
	$h_i \sim \text{gamma}(\alpha, \beta)$
	$\alpha \sim \text{gamma}(1,1)$
	$\beta \sim \text{gamma}(1,1)$
5. Negative Binomial 1 as Gamma-Poisson (Greene, 2008)	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon \times h_i)$
	$h_{i,t} \sim \text{gamma}(\epsilon \times \theta, \epsilon \times \theta)$
	$\epsilon \sim \text{gamma}(1,1)$
	$\theta \sim \text{uniform}(0, 250)$
6. Negative Binomial 2 as Gamma-Poisson (Greene, 2008)	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times \epsilon \times h_{it})$
	$h_{i,t} \sim \text{gamma}(\theta, \theta)$
	$\epsilon \sim \text{gamma}(1,1)$
	$\theta \sim \text{uniform}(0, 250)$
7. Lognormal with individual covariates ($\mathbf{X}$) and heterogeneity (σ)	$c_{i,t} \sim \text{Poisson}(a_{i,t} \times h_i)$
	$h_i \sim \text{lognormal}(\boldsymbol{\beta X}, \sigma^2)$
	$\boldsymbol{\beta} \sim \text{normal}(0, 10)$
	$\sigma \sim \text{gamma}(1,1)$

Table 2. Mean difference between the medians of the posterior distributions and truth and parameter coverage (in parentheses) for estimates of apparent survival (ϕ), availability for encounter given $a_{\text{i,t-1}} = 0$ (γ_1), availability for encounter given $a_{\text{i,t-1}} = 1$ (γ_2), primary occasion detection probability (p [CJS] or p^* [RD]), and the expected number of encounters per individual (ϵ) from 250 simulated capture-mark-recapture datasets analyzed using Cormack-Jolly-Seber (CJS; Cormack 1964; Jolly 1965; Seber 1965), robust design (RD; Kendall et al. 1997), zero-inflated Poisson (ZIP; this study), and zero-inflated Gamma-Poisson (ZIGP; this study) capture-recapture models.

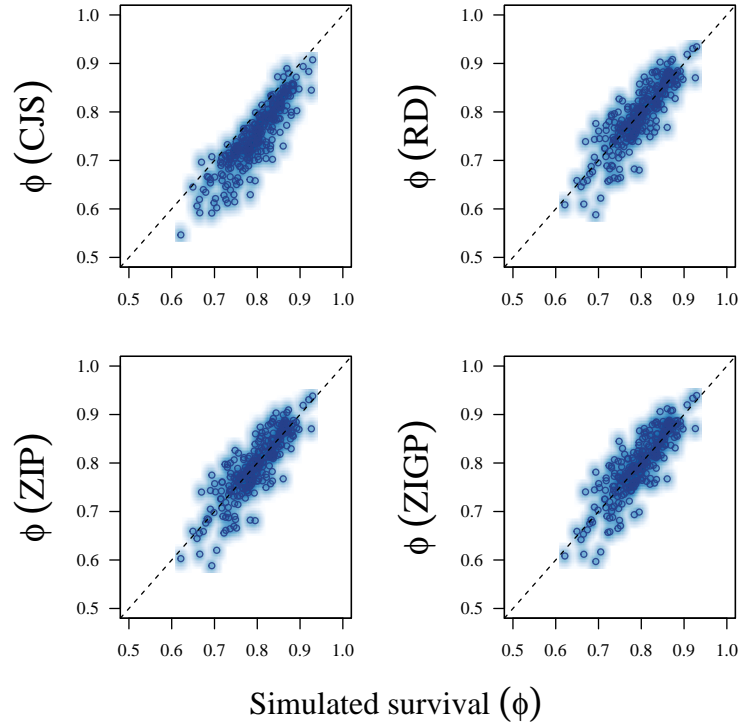| Parameter	CJS	RD	ZIP	ZIGP
ϕ	-0.047 (0.464)	-0.003 (0.940)	-0.002 (0.948)	0.001 (0.948)
γ_1	-	0.018 (0.956)	0.015 (0.964)	0.019 (0.976)
γ_2	-	-0.020 (0.892)	-0.013 (0.896)	0.006 (0.936)
p (CJS) or p^* (RD)	-0.306 (0.004)	0.010 (0.832)	-	-
ϵ	-	-	0.078 (0.764)	0.002 (0.928)

Figure 1. Scatter and density plots of the medians of posterior distributions for apparent survival relative to truth (ϕ) from Cormack-Jolly-Seber (CJS; upper left), robust design (RD; upper right), zero-inflated Poisson (ZIP, lower left), and zero-inflated gamma-Poisson with individual heterogeneity (ZIGP; lower right), capture-mark-reencounter models used to analyze 250 simulated capture-mark-reencounter datasets.
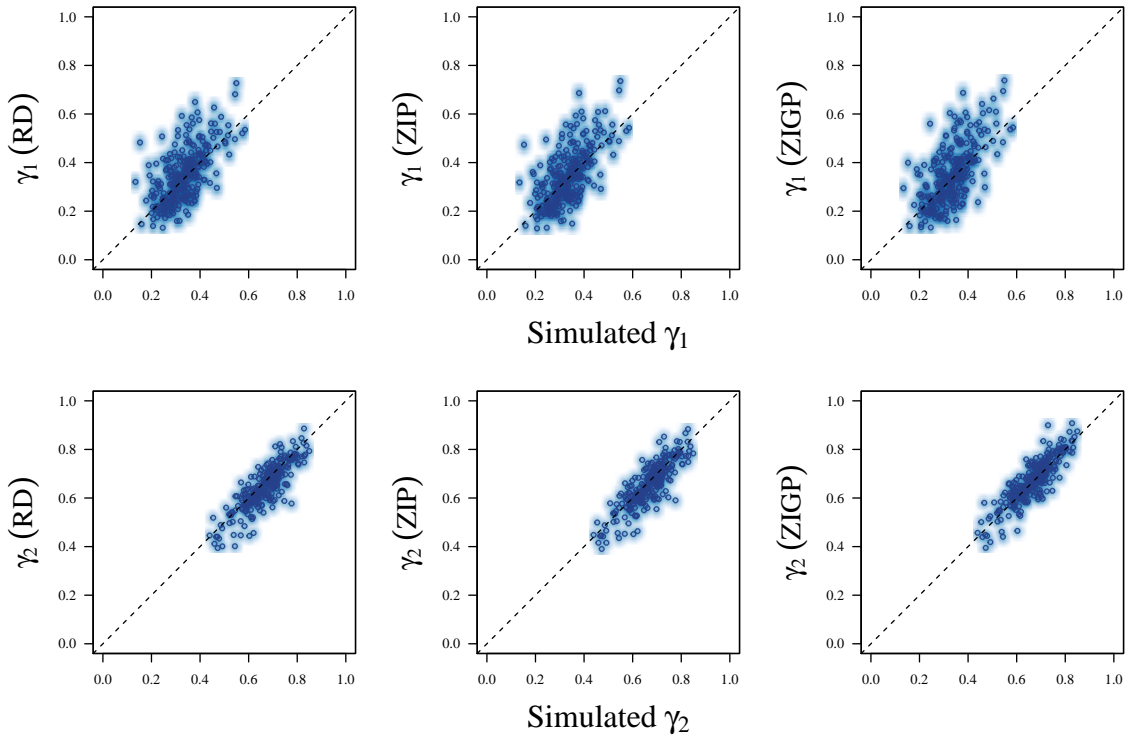
Figure 2. Scatter and density plots of the medians of posterior distributions for availability for encounter relative to truth (γ) from robust design (RD; left), zero-inflated Poisson (ZIP, center), and zero-inflated gamma-Poisson with individual heterogeneity (ZIGP; right), capture-mark-reencounter models used to analyze 250 simulated capture-mark-reencounter datasets.
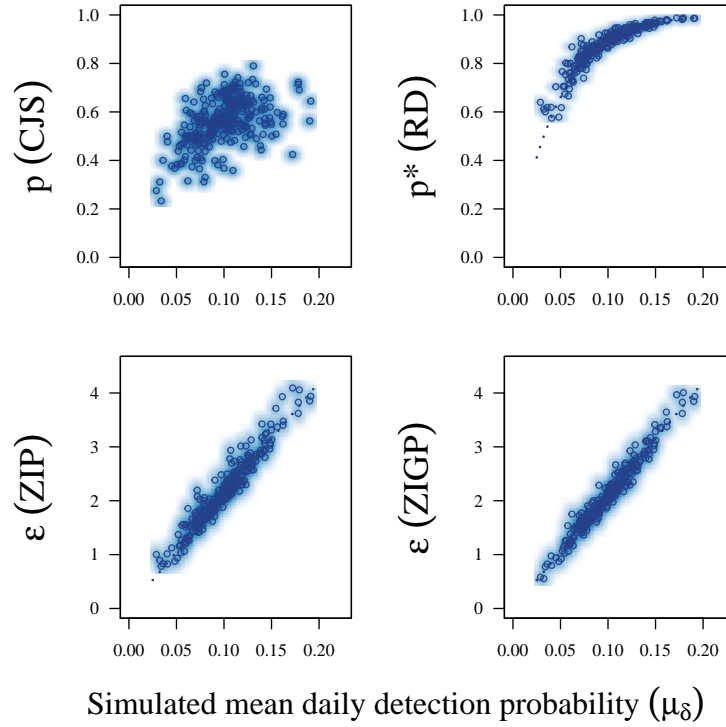
Simulated mean daily detection probability (μ_δ)

Figure 3. Scatter and density plots of the medians of posterior distributions for primary occasion detection probability (p) or the expected number of encounters per individual from Cormack-Jolly-Seber (CJS; upper left), robust design (RD; upper right), zero-inflated Poisson (ZIP, lower left), and zero-inflated Poisson with individual heterogeneity (ZIGP; lower right), capture-mark-reencounter models used to analyze 250 simulated capture-mark-reencounter datasets.
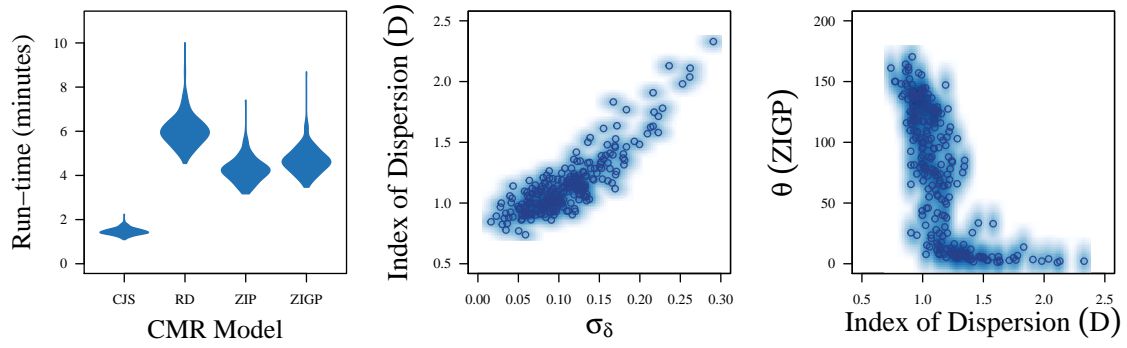
Figure 4. Violin plots of model run times across 250 simulations for Cormack-Jolly-Seber (CJS; Cormack 1964; Jolly 1965; Seber 1965), robust design (RD; Kendall et al. 1995, 1997), zero-inflated Poisson (ZIP; this study) and zero-inflated gamma-Poisson (ZIGP; this study) capture-mark-recapture models (left), scatter plots of the index of dispersion (D; Var(C)/Mean(C)) for the capture-mark-reencounter count data relative to the simulated heterogeneity in detection probability among individuals (σ_δ), and scatterplots of the mean of posterior distributions of the overdispersion parameter (θ) regressed against the index of dispersion for each capture-mark-recapture dataset.