

Citation:

Harris AJ, Wilson DR, Graham BA, Mennill DJ (2016) Estimating repertoire size in a songbird: a comparison of three techniques. *Bioacoustics*, 25: 211–224. doi: 10.1080/09524622.2016.1138416

Estimating repertoire size in a songbird: A comparison of three techniques

Alexander J. Harris^a, David R. Wilson^{a,b}, Brendan A. Graham^a, and Daniel J. Mennill^{a,*}

^aDepartment of Biological Sciences, University of Windsor, 401

Sunset Ave, Windsor, ON, Canada N9B 3P4

^bDepartment of Psychology, Memorial University of Newfoundland,

232 Elizabeth Avenue, St. John's, NL, Canada A1B 3X9

*Corresponding author. Email: dmennill@uwindsor.ca

Running head: Comparing repertoire size estimation techniques

1 **ABSTRACT**

2 Many animals produce multiple types of breeding vocalizations that, together, constitute a
3 vocal repertoire. In some species, the size of an individual's repertoire is important because it
4 correlates with brain size, territory size, or social behaviour. Quantifying repertoire size is
5 challenging because the long recordings needed to sample a repertoire comprehensively are
6 difficult to obtain and analyze. The most basic quantification technique is simple enumeration,
7 where one counts unique vocalization types until no new types are detected. Alternative
8 techniques estimate repertoire size from subsamples, but these techniques are useful only if
9 they are accurate. Using 12 years of acoustic data from a population of rufous-and-white wrens
10 in Costa Rica, we used simple enumeration to measure the repertoire size for 40 males. We
11 then compared these to the estimates generated by three estimation techniques: Curve Fitting,
12 Capture-Recapture, and a new technique based on the Coupon Collector's Problem. To
13 understand how sampling effort affects the accuracy and precision of estimates, we applied
14 each technique to six different-sized subsets of data per male. When averaged across subset
15 sizes, the Capture-Recapture and Coupon Collector techniques showed the highest accuracy,
16 whereas the Curve Fitting technique underestimated repertoire size. Precision (the average
17 absolute difference between the estimated and true repertoire size) was significantly better for
18 the Capture-Recapture technique than the Coupon Collector and Curve Fitting techniques. Both
19 accuracy and precision improved as subset size increased. We conclude that Capture-Recapture
20 is the best technique for estimating the sizes of small repertoires.

21 *Key words:* Capture-Recapture, Coupon Collector's problem, Curve Fitting, repertoire size,
22 simple enumeration, vocal repertoire

23 INTRODUCTION

24 Variation in vocal characteristics is associated with fitness in many species. For example,
25 structural variation in vocalizations can signal fighting ability and aggression (Linhart,
26 Slabbekoorn, and Fuchs, 2012), facilitate adaptive antipredator responses (Manser, 2013), and
27 enable animals to communicate effectively in the presence of variable background noise
28 (Slabbekoorn, 2013). Many animals have multiple types of breeding vocalizations that they can
29 produce, and, together, these constitute an animal's vocal repertoire. Repertoire sizes vary
30 considerably within species and populations (e.g. Peters et al., 2000), and this variation has
31 been correlated with reproductive success (e.g. Reid et al., 2004), territory size (e.g. Aweida,
32 1995), and cognitive abilities (e.g. Sewall, Soha, Peters, and Nowicki, 2013). Our understanding
33 of the adaptive significance of animal repertoires hinges on accurate and precise quantification
34 of repertoire size.

35 Determining an animal's repertoire size can be a challenging task. The most basic
36 technique is simple enumeration, which involves counting the number of unique types of
37 vocalizations that an individual produces. Ideally, an individual would be followed for its entire
38 lifetime to ensure that no vocalizations are missed. Because this is impractical, a rule must be
39 established to limit sampling effort to a practical level. The sampling effort required for simple
40 enumeration should reflect the effort typically needed to quantify an individual's entire
41 repertoire, based on previous findings that involve thorough recordings. If no previous findings
42 exist, then the sampling effort should be high enough that the researcher obtains many new
43 recordings without detecting any new vocalization types. The amount of effort required to
44 quantify repertoire size using simple enumeration is influenced by the size of the animal's

45 repertoire, the pattern with which the animal selects its vocalizations, the frequency with which
46 the animal vocalizes, and whether an animal is a closed-ended learner (i.e. all songs are learned
47 early in life and adult repertoire size is fixed) or an open-ended learner (i.e. songs continue to
48 be learned throughout life). Simple enumeration can work well for species with small repertoire
49 sizes, species that cycle through their entire repertoire cyclically, and species that vocalize often
50 (Botero et al., 2008). Simple enumeration requires much greater effort for species with larger
51 repertoires, species that choose different types of vocalizations with different probabilities or a
52 broader range of probabilities, and species that vocalize rarely.

53 Several estimation techniques have been developed to reduce the amount of effort
54 required to obtain an accurate measure of repertoire size. Two common techniques are Curve
55 Fitting and Capture-Recapture. The Curve Fitting technique uses the formula described by
56 Wildenthal (1965) to fit a line of best fit to a small subset of data. The horizontal asymptote of
57 the line then becomes the repertoire size estimate. Curve Fitting has been used for repertoire
58 size estimation in several species (Derrickson, 1987, Botero et al., 2008).

59 The Capture-Recapture technique involves a different approach that is based on a
60 comparison of the number of unique types of vocalizations recorded during two or more
61 sampling occasions. The proportion of vocalization types from an initial sample that are
62 observed again in subsequent samples is then used to estimate total repertoire size
63 (Baillargeon & Rivest, 2007). Capture-Recapture has been popular for estimating the sizes of
64 populations in ecological studies, but proves equally useful for estimating animals' repertoire
65 sizes (Garamszegi et al., 2002, Garamszegi et al., 2005).

66 Previous studies on the accuracy of Curve Fitting and Capture-Recapture techniques
67 have yielded inconsistent findings. Garamszegi et al. (2005) demonstrated that Capture-
68 Recapture could accurately estimate a bird's syllable repertoire size using only 15 songs. The
69 method was especially useful for species with large repertoires and heterogeneous selection
70 probabilities (Garamszegi et al. 2005). In another study that focused on species with large
71 repertoire sizes (≥ 160 element types), Botero et al. (2008) found that Capture-Recapture and
72 Curve Fitting were both inaccurate when the sample size was small, and that they only became
73 accurate when the sample size was so large that simple enumeration was also feasible (Botero
74 et al. 2008).

75 A new estimation technique based on the Coupon Collector's Problem (Erdős and Rényi,
76 1961; Feller, 1968; Dawkins, 1991) was recently debuted by Kershenbaum, Freeberg, and
77 Gammon (2015). The Coupon Collector's Problem describes a situation in which all items in a
78 set must be collected, and where sampling occurs with replacement. Under this model, the
79 initial items are collected rapidly, and the last few items take much more extensive sampling to
80 acquire. This situation has obvious parallels to the sampling of an animal's vocal repertoire,
81 particularly when the animals select their vocalization types at random (Kershenbaum et al.,
82 2015). Observed repertoire size grows rapidly at the beginning of sampling, but then tapers off
83 as more of the repertoire is sampled, until it plateaus when the entire repertoire has been
84 sampled. Using a modification of the Coupon Collector's Problem that accounts for unequal
85 probabilities of each song type (i.e. heterogeneous selection probability), Kershenbaum et al.
86 (2015) showed that this technique is a more accurate predictor of repertoire size than other
87 estimation techniques for species with heterogeneous selection probability. Their study

88 estimated repertoire sizes at the population level, rather than the individual level, and so
89 whether the Coupon Collector technique provides accurate estimates of the repertoire sizes of
90 individual animals remains to be studied.

91 In this study, we compare repertoire size estimation techniques by analyzing historical
92 repertoire size data from a population of rufous-and-white wrens (*Thryophilus rufalbus*). We
93 compare the Curve Fitting, Capture-Recapture, and Coupon Collector techniques in terms of
94 their ability to produce repertoire size estimates that match with the results from extensive
95 simple enumeration. Rufous-and-white wrens are neotropical songbirds found in forests
96 throughout western Central America and northwestern South America. Males of this species
97 are closed-ended learners that sing one song type repeatedly before switching to a different
98 song type, and may cycle through the same song types many times before singing their entire
99 repertoire (i.e. an eventual variety, non-cyclic singing style; Mennill & Vehrencamp 2005,
100 Hennin et al. 2009). When song type switches occur, certain song types are selected more often
101 than others, giving a heterogeneous selection probability to each song type in their repertoire
102 (unpublished data). Repertoire estimation techniques are thought to perform poorly when
103 animals are undersampled (Derrickson 1987) or when they do not select song types with equal
104 probability (Kroodsma 1982); this makes rufous-and-white wrens an interesting test case for
105 studying these three estimation techniques.

106 Our first goal was to determine the repertoire sizes of male rufous-and-white wrens
107 using twelve years of historical data collected in the field in Costa Rica. Many of our study
108 animals have been recorded extensively, and we could quantify their repertoire size with
109 confidence using simple enumeration. Our second goal was to compare the accuracy and

110 precision of repertoire size estimations from the Curve Fitting, Capture-Recapture, and Coupon
111 Collector techniques. We applied these techniques to different-sized subsets of our data and
112 compared the repertoire size estimates to the repertoire size we determined through simple
113 enumeration, which we used as a proxy for the animals' true repertoire sizes.

114 **METHODS**

115 **Recording Vocal Repertoires**

116 Data were collected at Sector Santa Rosa, Area de Conservación Guanacaste, Costa Rica
117 (10°40' N, 85°30' W), where our research group has been conducting a long-term study of
118 communication behaviour in a colour-banded population of rufous-and-white wrens since
119 2003. We analyzed data from 40 male wrens that we recorded during 1 to 7 successive
120 breeding seasons (average \pm SE: 3.7 ± 0.2) between 2003 and 2014. Birds were recorded
121 between March and July of each year, coinciding with the onset of the breeding season of this
122 species, when male vocal output reaches its peak (Topp & Mennill 2008). Birds were captured
123 in their territories using mist nets and then banded with a unique combination of three
124 coloured leg bands and a metal band to facilitate identification in the field. Rufous-and-white
125 wrens are renowned for their vocal duets (Mennill & Vehrencamp 2008, Kovach et al. 2014),
126 but we focused the current analyses on the vocalizations produced by males (both songs
127 produced as solos and as contributions to duets), given their high song output and our
128 extensive sampling of their songs (Mennill & Vehrencamp 2005; Topp & Mennill 2008).

129 **Analysis of Field Recordings**

130 We collected two types of field recordings: focal recordings and automated recordings.
131 Focal recordings involved a recordist following a male through his territory at distances of 10 to
132 30m, dictating the bird's identity after each song. All focal recordings were collected between
133 0445 h and 1100 h. Focal recordings were collected with a shotgun microphone (Sennheiser
134 MKH70 or ME67) and a solid-state digital recorder (Marantz PMD660 or PMD670; 22,050 Hz
135 sampling rate, 16-bit encoding accuracy, WAVE format). Focal recordings were collected every
136 year between 2003 and 2014, and they comprise the majority of recordings in this analysis
137 (approximately 60%).

138 To complement focal recordings, and to sample birds' repertoires over longer periods
139 than was possible with focal recordings, we collected automated recordings with three
140 different types of equipment, all used to sample birds' songs at times when focal recordists
141 were not present. (1) Microphone array recordings were collected in 2003 and 2004 by placing
142 an array of eight stationary omni-directional microphones throughout birds' territories
143 (sampling frequency: 22,050 Hz; full equipment details in Mennill et al., 2006). (2) Automated
144 recorders consisting of elevated omni-directional microphones (Sennheiser ME62) and solid-
145 state digital recorders (Marantz PMD670) were placed near the centre of the focal pair's
146 territory in 2007 through 2010 (sampling frequency: 44,100 Hz; full equipment details in
147 Mennill, 2014). (3) Automated Song Meter recorders (model: SM2-GPS, Wildlife Acoustics Inc.,
148 Concord, Massachusetts, USA) were placed in the centre of a pair's territory in 2011-2014,
149 usually within 10m of the focal pair's nest (sampling frequency: 22,050 Hz; full equipment
150 details in Mennill et al. 2012). We confirmed the identities of the birds in these unattended,
151 automated recordings by ensuring that the song types matched between the focal recordings

152 and the automated recordings; in all cases the songs recorded with the automated recorders
153 unambiguously matched with the songs in the focal recordings of the known male from the
154 same area. We distinguished between the voices of males versus females following previously
155 established criteria (see Mennill and Vehrencamp, 2005). Our ongoing field studies involve re-
156 sighting the birds throughout the field season to monitor their breeding behaviour, and we
157 ensured that focal animals were located in the same territory before and after automated
158 recordings were collected. Given that our study birds have large breeding territories (territory
159 sizes range from $5678 \pm 548 \text{ m}^2$ to $13497 \pm 1043 \text{ m}^2$; Osmun and Mennill, 2011, Mennill and
160 Vehrencamp, 2008), with substantial undefended spaces between adjacent territories (Osmun
161 and Mennill 2011), our automated recorders placed centrally within birds' territories recorded
162 only the target individuals. Any songs produced by rare territorial intruders were readily
163 distinguished from the resident birds by cross-referencing repertoire data of neighbouring
164 animals; even though song types are shared between individuals (Mennill & Vehrencamp 2005),
165 shared song types have individually distinctive characteristics.

166 **Assigning Songs to Song Types**

167 Rufous-and-white Wrens have vocal repertoires of songs, where each song type is
168 readily classified into different song types based on the visual and aural characteristics of the
169 three sections of their song: the introductory syllables, trill notes, and terminal syllables (as in
170 Mennill & Vehrencamp 2005; Barker 2008). Following previous work by Barker (2008), songs
171 were classified manually into types by comparing structural characteristics such as syllable
172 length, minimum and maximum frequencies, frequency of maximum amplitude, bandwidth,
173 and inter-syllable interval for the three song sections. In an analysis of song type categorization

174 that relied on discriminant analysis with cross-validation, Barker (2008) showed that fine
175 structural measurements are useful for accurately distinguishing different song types.

176 We annotated the audio files from all focal and automated recordings in SYRINX-PC
177 sound-analysis software (J. Burt, Seattle, Washington, USA). We annotated each song and
178 recorded its song type, manually comparing each song to a library of all previously recorded
179 song types from that animal. When a bird produced a song that had a different song type from
180 the previous song, we counted it as a song type switch. We determined the repertoire size of
181 each bird from the total number of song types recorded throughout the entire study for that
182 bird. Using these data, we constructed accumulation curves that showed the number of song
183 type switches sampled on the x-axis versus the number of unique song types detected on the y-
184 axis for each bird (Figure 1). Rather than using the total number of songs recorded, we used
185 song type switches as the unit of interest when calculating repertoire size (as in other studies,
186 for example, Valderrama et al. 2008; Sosa-López & Mennill 2014a, 2014b). We did this because
187 rufous-and-white wrens sing with eventual variety, repeating a given song type, on average,
188 eleven times before switching to a new song type (Mennill & Vehrencamp, 2005). Indeed, an
189 animal may sing a specific song type more than 100 times in a row before switching to a new
190 song type, leading to large plateaus in song type collection if sampling effort is measured
191 relative to number of songs sung instead of number of song type switches. Within these long
192 bouts of repeated songs, the song type of subsequent songs are not independent. For this
193 reason, we treated song type switches as our unit of analysis.

194 We used simple enumeration to measure the actual repertoire size of each rufous-and-
195 white wren because individuals used in this study had been recorded extensively (see Results).

196 This estimate was used as the benchmark to which the other three techniques were compared.
197 Only individuals with 150 or more recorded song type switches were used in the analysis. We
198 chose this number because 95% of the individuals had no new song types discovered after 150
199 song type switches using simple enumeration.

200 **Repertoire Size Estimation**

201 To determine the effect of sampling effort on the accuracy of each estimation
202 technique, we created subsets of the data for each bird, using the first 25, 50, 75, 100, 125, and
203 150 song type switches recorded from each individual. This allowed us to examine the
204 estimates produced by each technique from different amounts of sampling effort. We used R (R
205 Core Team, 2014) to generate data subsets and to generate all repertoire size estimates. The
206 raw data and the relevant R code are included in the online supplementary material.

207 For the Curve Fitting technique, we generated prediction curves for each possible
208 repertoire size between 1 and 30 song types (i.e. a range that encompassed repertoire sizes we
209 have encountered in our population in the last 12 years). We used the formula presented in
210 Wildenthal (1965):

$$211 \quad n = N (1 - e^{-T/N})$$

212 where n is the number of unique song types expected in a sample containing T song type
213 switches; N is the assumed repertoire size. Thus, for each possible repertoire size between 1
214 and 30 song types, we generated a unique curve with an asymptote at that value. We applied
215 an iterative process in which we generated a predictive model for each possible repertoire size,

216 and then assessed the fit of each model by comparing it to the observed data using a least
217 squares technique. Specifically, for each subset size and for each male, we selected the model
218 that generated the smallest value when the absolute differences between the predicted and
219 observed values were summed across all song type switches. The N from this model became
220 the best estimate of repertoire size.

221 For the Capture-Recapture technique, we used *Rcapture* (R package; Rivest &
222 Baillargeon, 2014) to estimate repertoire size. For each combination of male and subset size,
223 we created a capture history that indicated which song types were captured during which
224 capture occasions (0 = not captured; 1 = captured). Following Garamszegi et al. (2005) and
225 Botero et al. (2008), we defined a capture occasion as 5 song type switches, which divided
226 evenly into all of our subset sizes. Our Capture-Recapture models were based on a closed
227 population, since our preliminary analyses suggest that repertoire size does not change
228 throughout an adult's lifetime in this species (i.e. Rufous-and-white Wrens are closed-ended
229 learners; Mennill & Vehrencamp 2015; DJM unpublished data). *Rcapture* can incorporate
230 several different sources of variation that can each affect capture probabilities (Baillargeon &
231 Rivest, 2007). We used Darroch's M_h model, which allows the probability of capture to vary
232 among units (Darroch et al. 1993). This model thereby accounts for the possibility of common
233 and rare song types when predicting repertoire size.

234 The Coupon Collector's Problem is based on the idea of collecting a set of coupons that
235 are hidden in cereal boxes (Dawkins 1991; Feller 1968). If there are N different coupons, it
236 estimates the probability of collecting exactly i different coupons after purchasing m cereal
237 boxes. The coupons are drawn at random and with replacement. For our study, we used the

238 Coupon Collector's Problem to estimate the probability of observing i of N different song types
239 after sampling m song type switches. We implemented the Coupon Collector's Problem using a
240 Monte Carlo simulation. For each possible repertoire size (N), and for each possible number of
241 song type switches (m), we drew 100,000 independent samples. Each sample contained m
242 songs and was drawn at random and with replacement from the repertoire of N song types. As
243 in Kershenbaum et al. (2015), we modified the Coupon Collector's Problem to allow for unequal
244 probabilities of song type selection. We set the probability of selecting each song type based on
245 a Zipfian distribution, which has been used in previous studies to model the frequency of words
246 in human languages, as well as the frequency of song types in avian vocal repertoires (Zipf
247 1949; Lemon & Chatfield 1973). Probabilities are calculated by the formula:

$$248 \quad p(k; s, N) = \frac{1/k^s}{\sum_{n=1}^N (1/n^s)}$$

249 where $p(k; s, N)$ is the probability of selecting the k^{th} most common song type from a repertoire
250 of N song types; s is the absolute value of the slope of the regression of the frequency of each
251 song type on its corresponding rank, when plotted on a log-log scale. We used our raw data to
252 calculate s for each subset size included in our analyses (25, 50, 75, 100, 125, or 150 song type
253 switches). For each possible repertoire size (i.e. 1 to 30), and for each possible number of song
254 type switches (i.e. 1 to 150), we calculated the expected number of song types as the average
255 number of song types observed among the 100,000 samples. We used these values to create a
256 prediction curve for each repertoire size. As in our analysis of the Curve Fitting technique, we
257 assessed the fit of each prediction curve by comparing it to the observed data with a least

258 squares technique. The N from the model that minimized the least squares was selected as the
259 best estimate of repertoire size.

260 **Statistical Analysis**

261 We used a linear mixed-effects model in the R package *nlme* (Bates et al. 2015) to assess
262 the effects of estimation technique and subset size on the accuracy of repertoire size estimates.
263 We defined “accuracy” as the average difference between the repertoire size estimates
264 generated with a particular technique and the true repertoire sizes determined through simple
265 enumeration. In general, smaller deviations from zero indicated better accuracy; negative
266 values indicated that a method was underestimating the true repertoire size, whereas positive
267 values indicated that a method was overestimating the true repertoire size. We included the
268 differences as a dependent variable in our analysis, and the estimation technique (i.e. Curve
269 Fitting, Capture-Recapture, and the Coupon Collector technique), subset size (as a covariate),
270 and 2-way interaction as independent variables with fixed effects. We did not include an
271 intercept for the fixed effects because the hypothesized difference between the estimated and
272 observed repertoire sizes was zero. To facilitate the interpretation of model coefficients, we
273 centered subset size on zero. Bird identity was included as a subject variable with random
274 intercepts to account for repeated measurements from the same individuals. We fit the model
275 using restricted maximum likelihood estimation, and concluded that a particular estimation
276 technique was accurate if the difference between its repertoire size estimates and the true
277 repertoire sizes could not be distinguished statistically from zero.

278 We used a similar analysis to assess the effects of estimation technique and subset size
279 on the precision of repertoire size estimates. In this study, we consider precision to be a

280 measure of consistency in estimation. In estimating repertoire size, one might generate some
281 overestimates and some underestimates of true repertoire size, but an average value that
282 matches the true repertoire size; this is a situation with high accuracy, but low precision. We
283 defined “precision” as the average absolute difference between the repertoire size estimates
284 generated from a particular technique and the true repertoire sizes determined through simple
285 enumeration. Smaller differences in these absolute values would indicate more consistency in
286 the estimation of repertoire size, and therefore better precision. We again used a linear-mixed
287 effects model as in our analysis of accuracy (above). We compared precision among the three
288 estimation techniques using Tukey post-hoc comparisons, which we implemented in the R
289 package *multcomp* (Hothorn et al. 2008).

290 All tests were two-tailed, and results were considered significant when $p \leq 0.05$. Both
291 models complied with the parametric assumptions of linearity, homoscedasticity, and
292 normality, as revealed by visual inspection of residual plots.

293 **RESULTS**

294 **Enumerated Repertoire Size**

295 Simple enumeration showed that the 40 male rufous-and-white wrens produced an
296 average of 11.4 ± 0.3 song types each (mean \pm SE; range: 8 – 15 song types), which is in
297 accordance with a previous enumeration study of this species (Mennill and Vehrencamp, 2005).
298 These results were based on extensive recordings of each individual (e.g. Figure 1), containing
299 an average of 3619 ± 374 songs (mean \pm SE; range: 744 – 11691) and 447 ± 43 song type
300 switches (mean \pm SE; range: 154 – 1882).

301 Accuracy of Repertoire Size Estimates

302 Estimation technique had a significant effect on the accuracy of repertoire size
303 estimates (linear mixed-effects model: $F_{3,675} = 11.5$, $p < 0.001$; Fig. 2). The Capture-Recapture
304 technique generated repertoire size estimates that were not statistically different from animals'
305 true repertoire sizes ($t_{675} = -1.1$, $p = 0.283$; 95% CI for the difference: $-0.8 - 0.2$ songs types),
306 underestimating the true repertoire size by only 0.3 ± 0.3 song types (mean \pm SE; Table 1; Fig.
307 2). The Coupon Collector technique also generated repertoire size estimates that were not
308 significantly different from animals' true repertoire sizes ($t_{675} = 0.8$, $p = 0.444$; 95% CI: $-0.8 - 0.3$
309 song types), underestimating the true repertoire size by only 0.2 ± 0.3 song types. In contrast,
310 the Curve Fitting technique significantly underestimated repertoire size, with an average
311 repertoire size estimate that was 1.0 ± 0.3 song types below the true repertoire size ($t_{675} = -3.4$,
312 $p = 0.002$; 95% CI for the difference: $-1.5 - -0.4$ song types; Table 1; Fig. 2).

313 Subset size had a significant effect on the accuracy of repertoire size estimates, with
314 larger subset sizes producing more accurate estimates for all estimation techniques. This effect
315 was manifested through a significant interaction between estimation technique and subset size
316 (estimation technique: $F_{3,675} = 11.5$, $p < 0.001$; subset size: $F_{1,675} = 2.0$, $p < 0.001$; interaction:
317 $F_{2,675} = 14.8$, $p < 0.001$). Specifically, the Curve Fitting and Capture-Recapture techniques tended
318 to underestimate repertoire size more at smaller subset sizes than at larger subset sizes. In
319 contrast, the Coupon Collector technique tended to overestimate repertoire size more at
320 smaller subset sizes than at larger subset sizes (Table 1; Fig. 2).

321 Precision of Repertoire Size Estimates

322 The precision of repertoire size estimates was affected significantly by estimation
323 technique (linear mixed-effects model: $F_{3,675} = 13.3$, $p < 0.001$), subset size ($F_{1,675} = 201.0$, $p <$
324 0.001), and the 2-way interaction between them ($F_{2,675} = 6.0$, $p = 0.003$). The precision of the
325 Capture-Recapture technique was 0.9 ± 0.2 song types (mean \pm SE; 95% CI: 0.5 – 1.4 song
326 types), which was significantly better than the Coupon Collector technique (1.4 ± 0.2 song
327 types; 95% CI: 0.9 – 1.9 song types; Tukey post-hoc comparison: $Z = 6.3$, $p < 0.001$; Table 1), but
328 was statistically indistinguishable from the Curve Fitting technique (1.2 ± 0.2 song types; 95%
329 CI: 0.7 – 1.7 song types; Tukey post-hoc comparison: $Z = 2.2$, $p = 0.066$). The precision of the
330 Curve Fitting technique was statistically indistinguishable from the precision of the Coupon
331 Collector technique (Tukey post-hoc comparison: $Z = -1.5$, $p = 0.314$). Precision improved with
332 increasing subset size for all three techniques, although it improved more dramatically for the
333 Capture-Recapture and Coupon Collector techniques than it did for the Curve Fitting or
334 Capture-Recapture techniques (Table 1; Fig. 3).

335 DISCUSSION

336 Our comparison of three techniques for estimating song repertoire sizes of male rufous-
337 and-white wrens revealed that the Capture-Recapture and Coupon Collector techniques
338 produced more accurate estimates than the Curve Fitting technique, and that the Capture-
339 Recapture technique produced more precise estimates than the Coupon Collector and Curve
340 Fitting techniques. Both Capture-Recapture and Coupon Collector estimates were statistically
341 indistinguishable from actual repertoire size values based on simple enumeration, whereas

342 Curve Fitting estimates consistently underestimated the birds' repertoire sizes. Therefore, we
343 recommend using either Capture-Recapture or Coupon Collector estimation techniques for
344 generating accurate estimations of repertoire size, particularly for species with small or
345 medium sized repertoires, heterogeneous song type selection probability, and closed-ended
346 learning, like the rufous-and-white wren.

347 The Capture-Recapture technique had the best performance of the three techniques,
348 providing estimates that were statistically indistinguishable from our enumerated calculations
349 of repertoire size, and doing so even with a small sampling effort. The Capture-Recapture
350 technique estimated repertoire size to within 0.02 to 0.60 song types, and provided an
351 exceptionally accurate estimate of repertoire size with 100 or more song type switches (Figure
352 2). With subsets of just 25 song type switches, the repertoire size estimates derived from the
353 Capture-Recapture technique provided truer estimates than the Curve Fitting technique, as did
354 the Coupon Collector technique (Figure 2). Furthermore, the Capture-Recapture technique had
355 significantly better precision than the other two estimation techniques. Although precision was
356 similar for the three estimation techniques with small subsets of data, the Capture-Recapture
357 technique surpassed the precision of the other two techniques at higher sampling levels (Figure
358 3). Our conclusions are consistent with Garamszegi et al. (2005) who provided evidence that
359 Capture-Recapture is a compelling technique for estimating repertoire size.

360 The Coupon Collector technique is a newer estimation technique than the other two we
361 explore here. In the only other published study of the Coupon Collector technique,
362 Kershenbaum et al. (2015) found that this technique provided better estimates than the Curve
363 Fitting and Capture-Recapture techniques. Kershenbaum et al. (2015) generated estimates for

364 the very large repertoire sizes that exist among a population of animals, instead of the relatively
365 small repertoire sizes found within individuals. Our study is the first to assess the Coupon
366 Collector technique for estimating the repertoire sizes of individual animals. This technique was
367 the only technique that we explored here to over-estimate repertoire size, which occurred only
368 at our smallest sampling level (25 song type changes). At all higher sampling levels, the Coupon
369 Collector technique generated accurate estimates of repertoire size. Overall, the Coupon
370 Collector technique generated estimates with similarly high accuracy to the Capture-Recapture
371 technique, but with low accuracy at small sample sizes, and lower precision at all sample sizes.

372 The Curve Fitting technique produced estimates that underestimated repertoire size by
373 an entire song type. This underestimation likely arose due to uncommon song types present in
374 the repertoires of many rufous-and-white wrens. The Curve Fitting equation devised by
375 Wildenthal (1965) cannot account for uncommon song types because it is strongly affected by
376 the rapid presentation of common song types early in the sample. As sampling effort increased,
377 the Curve Fitting technique produced estimates with better accuracy and precision. Botero et
378 al. (2008) also explored the Curve Fitting technique for repertoire estimation and drew similar
379 conclusions that this technique underestimates repertoire size, especially when sampling effort
380 is small.

381 Many of our estimations resulted in under-estimates of repertoire size, including the
382 Curve Fitting technique estimations at all sampling levels, and the other two estimation
383 techniques at some sampling levels. Estimation techniques that under-estimate repertoire size
384 may still be well-suited for determining an individual's biologically relevant repertoire size. For
385 example, some birds in our study had song types that were only detected after thousands of

386 songs and hundreds of song type switches had already been recorded. Additionally, some song
387 types were very rare, and made up less than 0.1% of a bird's song production, occurring a few
388 times across multiple field seasons. Songs that are sung so infrequently that they require
389 extensive sampling to detect may have little impact on the bird's life history (Derrickson, 1987).
390 For example, in sedge warblers, repertoire size affects mate attraction (Buchanan and
391 Catchpole, 1997), so if females do not take the time to listen for rare song types, then rare song
392 types will have little to no impact on mate choice. Additionally, individuals may modify their
393 song type selection based on social contexts, and this could lead to a further decrease in an
394 individual's effective repertoire size. For example, Trillo and Vehrencamp (2005) found that
395 banded wrens modify their repertoire use in the presence of females to increase the
396 production of song types with specific acoustic characteristics and to song type match with
397 neighbouring males. Similarly, Hennin, *et al.* (2009) found that male rufous-and-white wrens
398 use a subset of their total repertoire when they are trying to attract a mate. Techniques that
399 consistently under-estimate repertoire size, as we have revealed for the Curve Fitting technique
400 here, may offer realistic estimates of how other birds assess an individual's repertoire size by
401 ignoring rare song types or song types that are not used in specific social contexts.

402 Overall, we found that the Capture-Recapture and Coupon Collector techniques provide
403 the most accurate estimates of repertoire size, and that the Capture-Recapture technique
404 provides the most precise estimates of repertoire size. The Curve Fitting technique did not
405 perform as well, tending to underestimate repertoire size to a statistically significant degree at
406 smaller sample sizes. Future research should explore the use of Capture-Recapture for
407 estimating actual repertoire size in other species with small repertoire sizes and heterogeneous

408 song type probabilities. Curve Fitting and Coupon Collector techniques may be useful for
409 estimating an individual's biologically relevant repertoire size in contexts where rare song types
410 have little to no impact.

411

412 **ACKNOWLEDGEMENTS**

413 We thank the staff of Sector Santa Rosa of the Guanacaste Conservation area for logistical
414 support. We thank A. Kirshebaum and an anonymous reviewer for ideas that improved the
415 manuscript. This research was supported by a Natural Sciences and Engineering Research
416 Council of Canada (NSERC) Undergraduate Summer Research Award to AJH; by an Ontario
417 Graduate Scholarship (OGS), a Queen Elizabeth II Graduate Scholarship in Science and
418 Technology (QEII-GSST), a Chapman Grant from the American Museum of Natural History, a
419 Student Research Grant from the Animal Behaviour Society, and an Alexander Wetmore
420 Research Award from the American Ornithologist Union to BAG; by an NSERC Postdoctoral
421 Fellowship to DRW; and by an NSERC Discovery Grant, an NSERC Accelerator Grant, and two
422 NSERC Research Tools and Instruments Grants to DJM. Further support was provided by the
423 Canadian Foundation for Innovation, the Government of Ontario, and the University of Windsor
424 to DJM.

425

426 **LITERATURE CITED**

- 427 Aweida, M. K. (1995). Repertoires, territory size, and mate attraction in western meadowlarks.
428 *The Condor* 97:1080-1083.
- 429 Baillargeon, S. and L.-P. Rivest (2007). Rcapture: loglinear models for Capture-Recapture in R.
430 *Journal of Statistical Software* 19:1-31.

431 Barker, N. K. (2008). Effective communication in tropical forests: song transmission and the
432 singing behaviour of rufous-and-white wrens (*Thryophilus rufalbus*) (Master's thesis).
433 University of Windsor.

434 Botero, C. A., A. E. Mudge, A. M. Koltz, W. M. Hochachka, and S. L. Vehrencamp (2008). How
435 reliable are the methods for estimating repertoire size? *Ethology* 114:1227-1238.

436 Buchanan, K. L., and C. K. Catchpole (1997). Female choice in the sedge warbler *Acrocephalus*
437 *schoenobaenus*: multiple cues from song and territory quality. *Proceedings of the Royal*
438 *Society of London B* 24:521-526.

439 Darroch, J. N., Fienberg, S. E., Glonek, G. F. V, and B. W. Junker (1993). A three-sample multiple-
440 recapture approach to census population estimation with heterogeneous catchability.
441 *Journal of the American Statistical Association* 88:1137-1148.

442 Dawkins, B. (1991) Siobhan's problem: the Coupon Collector revisited. *The American Statistician*
443 45:76-82.

444 Derrickson, K. C. (1987). Yearly and situational changes in the estimate of repertoire size in
445 northern mockingbirds (*Mimus polyglottos*). *The Auk* 104:198-207.

446 Erdős, P. and A. Rényi (1961). On a classical problem of probability theory. *Magyar Tudományos*
447 *Akadémia Értesítője* 6:215-220.

448 Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (vol. I, 3rd ed.),
449 New York: John Wiley.

450 Garamszegi, L. Z., T. J. S. Balsby, B. D. Bell, M. Borowiec, B. E. Byers, T. Draganoiu M. Eens, W.
451 Forstmeier, P. Galeotti, D. Gil, L. Gorissen, P. Hansen, H. M Lampe, S. Leitner, J.
452 Lontkowski, L. Nagle, E. Nemeth, R. Pinxten, J.-M. Rossi, N. Saino, A. Tanvez, R. Titus, J.
453 Török, E. Van Duyse, and A. P. Møller (2005). Estimating the complexity of bird song by
454 using Capture-Recapture approaches from community ecology. *Behavioral Ecology and*
455 *Sociobiology* 57:305-317.

456 Garamszegi, L. Z., T. Boulinier, A. P. Møller, J. Török, G. Michl, and J. D. Nichols (2002). The
457 estimation of size and change in composition of avian repertoires. *Animal Behavior*
458 63:623-630.

459 Hennin, H. L., N. K. S. Barker, D. W. Bradley, and D. J. Mennill (2009). Bachelor and paired male
460 rufous –and-white wrens use different singing strategies. *Behavioral Ecology and*
461 *Sociobiology* 64:151-159.

462 Hothorn, T., Bretz, F. and P. Westfall (2008). Simultaneous inference in general parametric
463 models. *Biometrical Journal* 50:346-363.

464 Kershenbaum, A., T. M. Freeberg, and D. E. Gammon (2015). Estimating vocal repertoire size is
465 like collecting coupons: A theoretical framework with heterogeneity in signal
466 abundance. *Journal of Theoretical Biology* 373:1-11.

467 Kroodsma, D.E. (1982). Song repertoires: problems in their definition and use. In: Kroodsma DE,
468 Miller EH, editors. Acoustic Communication in Birds: Song Learning and its
469 Consequences. Academic Press; New York, New York, USA.

470 Lemon, R. E., and C. Chatfield (1973). Organization of song of rose-breasted grosbeaks. *Animal*
471 *Behaviour* 21: 28-44.

472 Linhart, P., H. Slabbekoorn, and R. Fuchs (2012). The communicative significance of song
473 frequency and song length in territorial chiffchaffs. *Behavioral Ecology* 23:1338-1347.

474 Manser, M. B. (2013). Semantic communication in vervet monkey and other animals. *Animal*
475 *Behavior* 86:491-496.

476 Mennill, D. J. (2014). Variation in the vocal behavior of common loons (*Gavia immer*): insights
477 from landscape-level recordings. *Waterbirds* 37:26-36.

478 Mennill, D.J., M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet (2012). Field test of an
479 affordable, portable, wireless microphone array for spatial monitoring of animal ecology
480 and behaviour. *Methods in Ecology and Evolution* 3:704-712.

481 Mennill, D. J., J. M. Burt, K. M. Fristrup, and S. L. Vehrencamp (2006). Accuracy of an acoustic
482 location system for monitoring the position of duetting tropical songbirds. *Journal of the*
483 *Acoustical Society of America* 119:2832-2839.

484 Mennill, D. J. and S. L. Vehrencamp (2005). Sex differences in the singing and duetting behavior
485 of neotropical rufous-and-white wrens (*Thryothorus rufalbus*). *Auk* 122:175-186.

486 Mennill, D. J. and S. L. Vehrencamp (2008). Context-dependent functions of avian duets
487 revealed through microphone array recordings and multi-speaker playback. *Current*
488 *Biology* 18:1314-1319.

489 Osmun, A. E., and D. J. Mennill (2011). Acoustic monitoring reveals congruent patterns of
490 territorial singing behaviour in male and female tropical wrens. *Ethology* 117:385-394.

491 Peters, S., W. A. Searcy, Beecher M. D., and S. Nowicki (2000). Geographic variation in the
492 organization of song sparrow repertoires. *Auk* 117:936-942.

493 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2015). nlme: linear and nonlinear
494 mixed effects models. R package version 3.1-122.

495 R Core Team (2014). R: a language and environment for statistical computing. R Foundation for
496 Statistical Computing, Vienna, Austria.

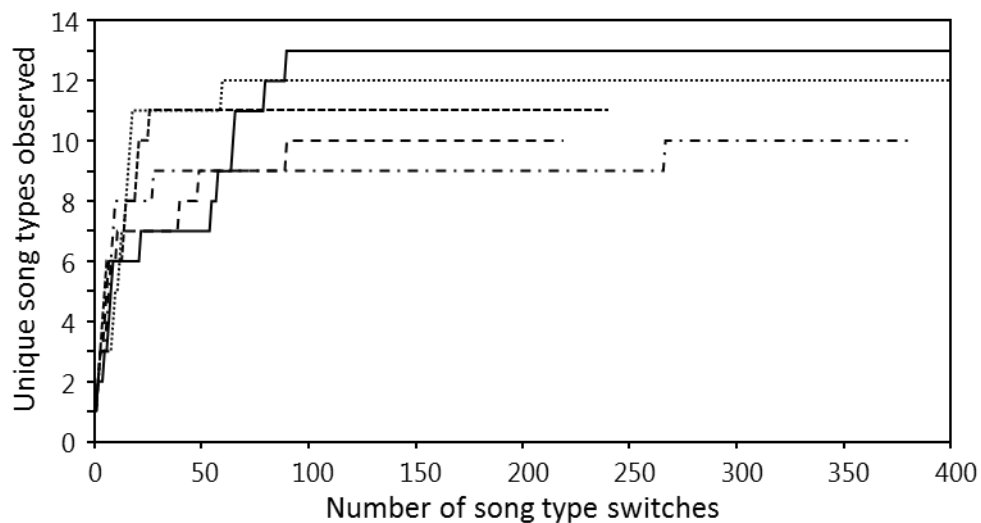
497 Reid, J. M., P. Arcese, A. L. E. V. Cassidy, S. M. Hiebert, J. N. M. Smith, P. K. Stoddard, A. B. Marr,
498 and L. F. Keller (2004). Song repertoire size predicts initial mating success in male song
499 sparrows, *Melospiza melodia*. *Animal Behavior* 68:1055-1063

500 Rivest, L.-P. and S. Baillargeon (2014). Rcapture: loglinear models for Capture-Recapture
501 experiments. R package version 1.4-2.

502 Sewall, K. B., J. A. Soha, S. Peters, and S. Nowicki (2013). Potential trade-off between vocal
503 ornamentation and spatial ability in a songbird. *Biology Letters* 9.

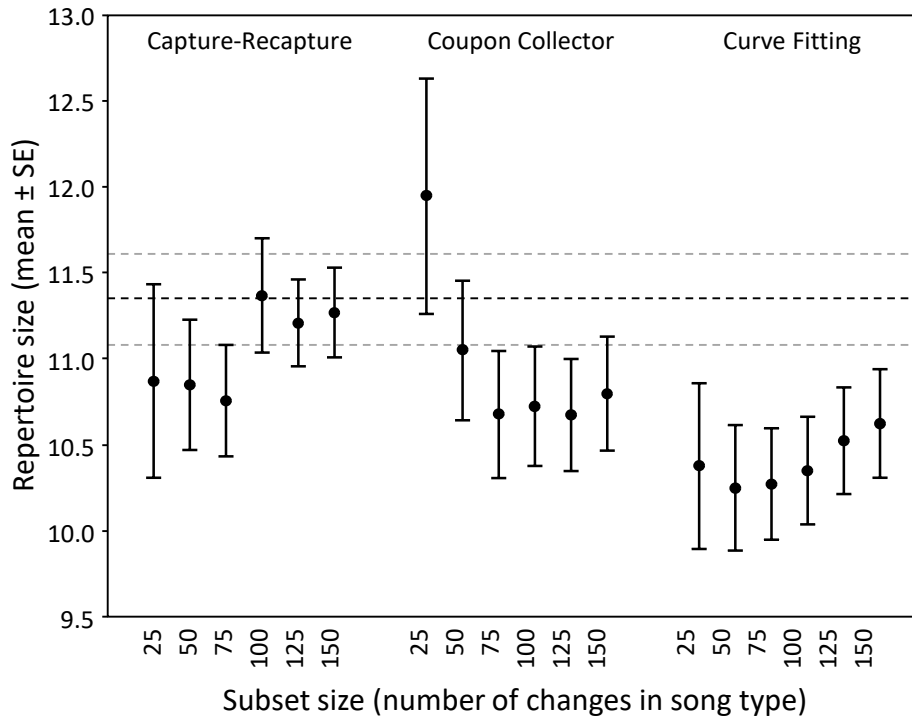
- 504 Slabbekoorn, H. (2013). Songs of the city: noise-dependent spectral plasticity in the acoustic
505 phenotype of urban birds. *Animal Behavior* 85:1089-1099.
- 506 Sosa-Lopez, J.R. and D. J. Mennill (2014a). The vocal behaviour of the Brown-throated Wren
507 (*Troglodytes brunneicollis*): Song structure, repertoires, sharing, syntax, and diel
508 variation. *Journal of Ornithology* 155:435-446.
- 509 Sosa-Lopez, J.R. and D. J. Mennill (2014b). Vocal behaviour of the island-endemic Cozumel
510 Wren (*Troglodytes aedon beani*): Song structure, repertoires, and song sharing. *Journal*
511 *of Ornithology* 155:337-346.
- 512 Trillo, P. A. and S. L. Vehrencamp (2005). Song types and their structural features are associated
513 with specific contexts in the banded wren. *Animal Behavior* 70:921-935.
- 514 Valderrama, S., Parra, J., Davila, N. and D.J. Mennill (2008). The vocal behavior of the critically-
515 endangered Niceforo's Wren (*Thryothorus nicefori*). *Auk* 125:395-401.
- 516 Wildenthal, J. L. (1965). Structure in primary song of the mockingbird (*Mimus polyglottos*). *The*
517 *Auk* 82:161-189.
- 518 Zipf, G. K. (1949) Human behavior and the principle of least effort. Addison-Wsley Press,
519 Cambridge, MA.
- 520

521 **Figures**



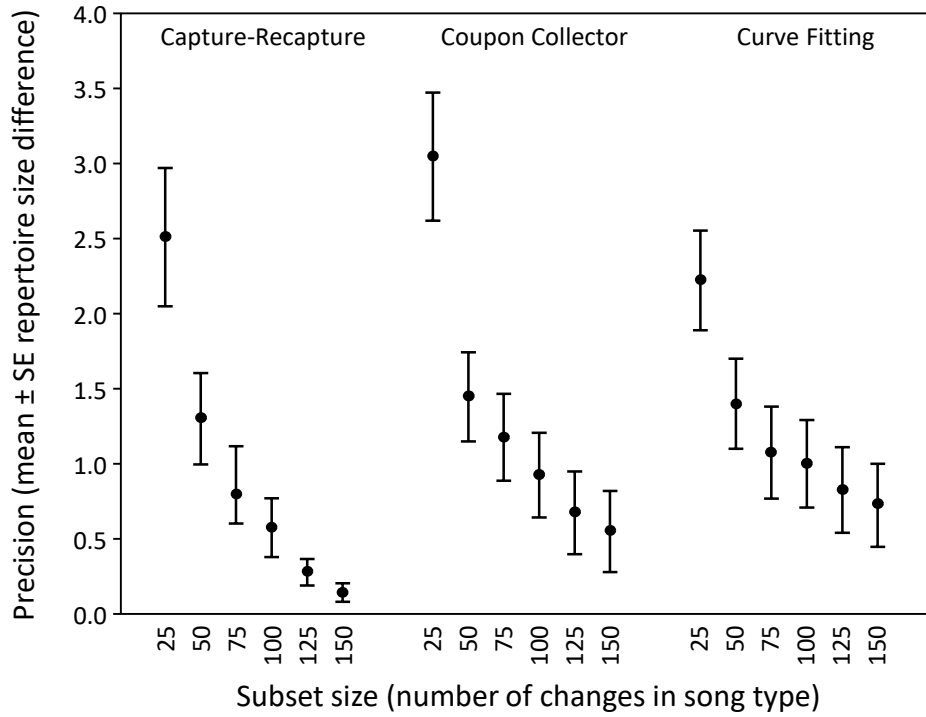
522
523 **Figure 1.** Simple enumeration data showing repertoire size estimates for five example male
524 rufous-and-white wrens. Sampling effort (number of song type switches recorded) is on the x-
525 axis and number of unique song types detected is on the y-axis. The large plateaus in the graph,
526 where the number of unique song types does not increase despite large increases in sampling
527 effort, suggest that a bird's repertoire has been sampled in its entirety. Note that in rare cases,
528 such as the lowest curve, unique songs are detected even after extensive sampling.

529



530

531 **Figure 2.** Estimated repertoire sizes from three different estimation techniques (Capture-
 532 recapture, Coupon Collector, and Curve Fitting techniques) for 40 male rufous-and-white
 533 wrens. For each of the three estimation techniques, the error bars show estimated repertoire
 534 sizes for each of six subset sizes: 25 (left), 50, 75, 100, 125, and 150 (right) song type switches.
 535 True repertoire sizes were measured through simple enumeration and are depicted by the
 536 hatched lines (mean = black hatched line; mean \pm SE = gray hatched lines). Accuracy is defined
 537 as the average difference between the repertoire size estimated with a particular technique
 538 and subset size and the true repertoire size determined through simple enumeration. Smaller
 539 differences indicate better accuracy.



540

541 **Figure 3.** Effects of estimation technique and subset size on the precision of repertoire size
 542 estimates for 40 male rufous-and-white wrens. For each estimation technique, we show the
 543 precision of repertoire size estimates derived from subsets of 25 (left), 50, 75, 100, 125, and
 544 150 (right) song type switches. Precision is defined as the average absolute difference between
 545 true repertoire size, as determined through simple enumeration, and the repertoire size
 546 estimated with a given technique and subset size; smaller absolute differences indicate better
 547 precision.

548 **Table 1.** Model coefficients from the analyses of the accuracy and precision of repertoire size
 549 estimates.

Dependent Variable	Parameter	Model Coefficient	SE
Accuracy	Capture-Recapture ¹	-0.298	0.277
	Curve Fitting ¹	-0.950	0.277
	Coupon Collector ¹	-0.213	0.277
	Subset size ²	0.004	0.002
	Curve fitting x subset size ³	-0.002	0.003
	Coupon collector x subset size ³	-0.017	0.003
Precision	Capture-Recapture ¹	0.946	0.240
	Curve Fitting ¹	1.208	0.240
	Coupon Collector ¹	1.379	0.240
	Subset size ²	-0.017	0.002
	Curve fitting x subset size ³	0.007	0.003
	Coupon collector x subset size ³	-0.002	0.003

550 ¹Model coefficients for the three estimation techniques indicate the average
 551 accuracy or precision of the technique (in song types), relative to zero, when all other
 552 variables are held constant.

553 ²Model coefficients for subset size indicates how much the dependent variable
 554 changes (in terms of song types) with each 1-unit change in subset size, when
 555 averaged across all techniques.

556 ³Model coefficients for the interaction terms indicate how much the dependent
 557 variable changes with each 1-unit change in subset size for a particular technique,
 558 relative to the amount of change observed for a 1-unit change in subset size for the
 559 reference category (i.e. Capture-Recapture).