

1 **A ROTATION TEST FOR BEHAVIOURAL POINT PROCESS DATA**

2 Running Headline: DeRuiter and Solow, Rotation Test for Behavioral Data

3 STACY L. DERUITER & ANDREW R. SOLOW

4 Woods Hole Oceanographic Institution

5 Woods Hole, MA

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24 Addresses for Correspondence:

25 Stacy DeRuiter

26 Biology Department, MS #50

27 Woods Hole Oceanographic Institution

28 Woods Hole, MA 02543

29 USA

30 Email: sderuiter@whoi.edu

31

32 Andrew Solow

33 Marine Policy Center, MS #41

34 Woods Hole Oceanographic Institution

35 Woods Hole, MA 02543

36 USA

37

38 Word Count: 3079

39 **Key Words:** behavioural point process, statistical analysis, time-series analysis

40 A common problem in animal behavior is determining whether the rate at which a certain
41 behavioural event occurs is affected by an environmental or other factor. In the example
42 considered later in this paper, the event is a vocalization by an individual sperm whale
43 and the factor is the operation or non-operation of an underwater sound source. A typical
44 experiment to test for such effects involves observing animals during control and
45 treatment periods and recording the times of the events that occur in each. In statistical
46 terminology, the data arising from such an experiment – the times at which events of a
47 specified type occur – represent a point process (Cox & Lewis 1978). Events in a point
48 process are treated as having no duration. Although this is not strictly correct for
49 behavioural events, the approximation is reasonable when the duration of events is small
50 in relation to the interval between them.

51 In some cases, under the null hypothesis of no treatment effect, behavioural
52 events can be assumed to follow a stationary Poisson process. Under this model, the
53 intervals between successive events are independent and, conditional on their number, the
54 events are uniformly distributed over the observation period. As described below, when
55 the Poisson assumption is valid, a statistical test to determine whether event rate changes
56 under treatment can be based on the binomial distribution. In many cases, however, the
57 Poisson model has been shown to be invalid for behavioural events. This is the case, for
58 example, when events occur in bouts (Slater & Lester 1982; Sibly et al. 1990; Haccou &
59 Meelis 1992). As illustrated below, when behavioural events do not follow a Poisson
60 process, the binomial test can give misleading results. A number of methods are
61 available to test whether a point process is Poisson based on the uniformity result
62 mentioned above (Stephens 1986). If a point process cannot be assumed to be Poisson,

63 one option is to use a test that is valid under a particular alternative to the Poisson model.
64 Unfortunately, while it is often easy to show that a point process is not Poisson, it can be
65 difficult to specify an appropriate alternative model. The purpose of this paper is to
66 describe and illustrate the use of a simple nonparametric method that can be used to
67 analyze behavioural point process data even if the process generating the data is
68 unknown.

69

70 **A Rotation Test**

71

72 Suppose that events are observed over the period $(0, T)$, and that the total times under
73 control and treatment conditions are T_C and T_T , respectively (with $T = T_C + T_T$). Assume
74 that, under control conditions, events follow a stationary Poisson process with rate λ_C
75 and that, under treatment conditions, events follow an independent Poisson process with
76 rate λ_T . Under this model, interest centers on testing the null hypothesis $H_o : \lambda_C = \lambda_T$ of
77 no treatment effect. Let the random variables N_C and N_T be the numbers of events
78 occurring under control and treatment conditions, respectively, and let $N = N_C + N_T$ be
79 the total number of events. Conditional on the observed value n of N , under H_o , N_C has
80 a binomial distribution with n trials and success probability T_C / T . The null hypothesis
81 can be rejected at significance level α if the observed value of N_C is below the lower
82 $(\alpha / 2)$ quantile or above the upper $(\alpha / 2)$ quantile of this binomial distribution.

83 Provided n is not too small and T_C / T is not too close to 0 or 1, the binomial distribution

84 can be approximated by a normal distribution with mean nT_c/T and variance

85 $nT_c T_T / T^2$, so that H_o can be rejected at approximate significance level α if:

86

87
$$\left| \frac{TN_c - nT_c}{\sqrt{nT_c T_T}} \right| > \Phi^{-1}(\alpha/2) \quad (1)$$

88

89 where $\Phi^{-1}(\alpha/2)$ is the upper $(\alpha/2)$ quantile of the standard normal distribution.

90

91 Behavioral events often exhibit clustering in time beyond what is expected under
92 a Poisson process. As illustrated below, the binomial test may fail in such cases because
93 the underlying randomization scheme – distributing n events at random over the
94 observation period – fails to capture internal structure in the events that is present even
95 under the null hypothesis. A randomization procedure that does preserve this internal
96 structure can be visualized in the following way. Transform the observation period into a
97 circle by joining its end to its beginning. This is sometimes referred to as imposing a
98 periodic boundary condition. Keeping the partition of the observation period into control
99 and treatment segments fixed, displace the events by the same random rotation. Let t_j be
100 the time of the j th event. Its time under this rotation scheme is given by:

101

102
$$\begin{aligned} t_j^* &= t_j + U & t_j + U < T \\ & t_j + U - T & t_j + U > T \end{aligned} \quad (2)$$

103

104 where U is a uniform random variate over the interval $(0, T)$. By displacing each event
105 by the same random angle, this procedure preserves the internal structure of the events
106 except at the beginning of the original observation period, where events originally near T
107 are now in proximity to events originally near 0. Provided n is not too small, the effect of
108 this concatenation is negligible. The test proceeds by approximating the distribution of
109 N_c under the null hypothesis from values produced by a large number of random
110 rotations. The null hypothesis is then rejected at significance level α if the observed
111 value of N_c lies below the lower or above the upper $\alpha/2$ quantile of this distribution.
112 This general approach was originally proposed by Harkness & Isham (1983) for testing
113 association between two two-dimensional point processes observed on a rectangle. The
114 test has been applied by Miller et al (2004a, b), but a detailed description of the test
115 including assessment of its validity and power has not previously been published.

116

117 To summarize, the steps involved in the rotation test are:

118

- 119 1. Express the data as a set of behavioral event times over the observation period $(0,$
120 $T)$.
- 121 2. Calculate the number N_c of events occurring during the control period.
- 122 3. Generate a rotated set of behavioural event times according to (6).
- 123 4. Calculate the number N_{c_rots} of events in the rotated set falling in the original
124 control time period.

125 5. Repeat steps 3-5 many times to obtain a distribution for $N_{c_{rot}}$ and assess
126 significance by comparing the the observed value of N_C to the quantiles of this
127 distribution.

128

129 **Test Performance**

130

131 In this section, we assess the performance of the binomial and rotation tests under three
132 point process models: the stationary Poisson process, the one-dimensional Thomas
133 process (Thomas 1949), and an exponential autoregressive (EAR) model (Lawrance &
134 Lewis 1979). The Thomas process is a classical model of clustering in point process
135 data, while the EAR model gives rise to clustering through positive autocorrelation in the
136 intervals between events. No claim is made that either of these models is necessarily
137 appropriate for a particular behavioural point process. Instead, they are used here as
138 plausible alternatives to the Poisson process.

139

140 In a Thomas process, initiating events follow a stationary Poisson process with
141 rate μ . Each initiating event gives rise to an additional number of offspring events. The
142 numbers of these offspring are independent Poisson random variables with mean θ . Let
143 s_o be the time of an initiating event and suppose that it gives rise to k offspring. The
144 times of these offspring are given by $s_j = s_o + \delta_j, j = 1, 2, \dots, k$, where $\delta_1, \delta_2, \dots, \delta_k$ are
145 independent random intervals with common distribution function F . The process consists
146 of the union of the initiating events and their offspring. The Thomas process is stationary
147 with overall rate $\mu(1 + \theta)$. However, it is over-dispersed in relation to the Poisson

148 process with the same rate. For example, for the Thomas process, the variance of the
149 number of events occurring in a unit interval is $\mu(1+3\theta+\theta^2)$ instead of $\mu(1+\theta)$ for the
150 Poisson process with the same overall rate.

151

152 In contrast to the Thomas process, which is a model of the event times, the EAR
153 process is a model for the intervals between events. Let $d_j = t_j - t_{j-1}$ be the interval
154 between events $j-1$ and j . Under the Poisson model, the intervals d_1, d_2, \dots are
155 independent exponential random variables. In contrast, under the EAR model, the
156 sequence of intervals follows the autoregressive process:

157

$$158 \quad d_j = \rho d_{j-1} + \varepsilon_j \quad (3)$$

159

160 where ε_j is equal to 0 with probability ρ and equal to an exponential random variable
161 with mean $1/\lambda$ with probability $1-\rho$. The EAR process is stationary with overall rate
162 λ and autocorrelation function $Corr(d_j, d_{j-h}) = \rho^h$. The positive dependence between
163 successive intervals gives rise to clustering of events.

164

165 Although it is possible to make some progress analytically, for the purpose of this
166 paper we present some results from a small simulation study. The goal of the first part of
167 this study was to assess the validity of the nominal significance levels of the binomial and
168 rotation tests under the three point process models outlined above. This involved
169 repeatedly simulating point process data from these models under the null hypothesis and

170 applying both tests at the nominal 0.05 significance level. For a valid test, the null
171 hypothesis should be rejected at a rate equal to the nominal significance level. In the
172 study described here, the observation period was taken to be the unit interval, with the
173 first half corresponding to the control period and the second half to the treatment period.
174 Results are presented in Table 1 for overall mean rates of 500 and 1000. For the Thomas
175 process, the parameter θ was fixed at 1 while for the EAR process the parameter ρ was
176 fixed at 0.5. Each entry in Table 1 was based on 1000 simulated data sets and each
177 rotation test was based on 1000 random rotations. In the case of the Thomas process, we
178 assumed that offspring events fell into the same sub-period as their initiating event. In
179 practical terms, this amounts to the assumption that the displacements between offspring
180 and initiating events are negligible in relation to the length of the periods of control and
181 treatment conditions.

182

183 Turning to Table 1, it is clear that the binomial test is invalid for point process
184 data generated by the Thomas and EAR processes. For these models, the estimated true
185 rate at which the null hypothesis is falsely rejected is well above the nominal significance
186 level. In contrast, the estimated true significance level for the rotation test is not
187 significantly different from the nominal level for all three point process models.

188

189 The goal of the second part of the simulation study was to assess the power of the
190 rotation test. Power is defined as the probability of rejecting the null hypothesis when the
191 alternative hypothesis is correct. This probability will depend on the nature and
192 magnitude of the departure from the null hypothesis, as well as on the amount of data.

193 As a rough guide, a test has good power if this probability is at least 0.8. The power
194 study was based on the same general simulation procedure outlined above except that, for
195 each of the point process models, the overall rate under treatment conditions was
196 increased by a multiplicative factor f over its value under control. For the Thomas
197 process, this was accomplished by increasing the rate μ of initiating events. As before,
198 let λ_C and λ_T be the rates under control and treatment conditions, respectively. For the
199 case here where the observation period is evenly divided between control and treatment,
200 the overall rate λ is simply the average of λ_C and λ_T . Throughout this power study,
201 this overall rate was held fixed by taking $\lambda_C = 2\lambda/(1+f)$ and $\lambda_T = f\lambda_C$.

202

203 The results of the power study are shown in Table 2. Results are presented for
204 overall rates 500 and 1000 with the parameter θ of the Thomas process fixed at 1, the
205 parameter ρ of the EAR process fixed at 0.5, and $f = 1.5, 2, \text{ and } 3$. . As before, each
206 entry in this table was based on 1000 simulated data sets and for each data set the rotation
207 test was based on 1000 random rotations. For the Poisson case, results are presented for
208 both the binomial test and the rotation test. In this case, the rotation test is less powerful
209 than the binomial test, although it achieves good power in most of the cases considered
210 here. As the binomial test is not valid for the Thomas and EAR processes, for these
211 processes results are presented only for the rotation test. The power of the rotation test is
212 quite similar for the two cluster processes. In general, the rotation test achieves good
213 power provided the magnitude of the treatment effect and the overall rate of events are
214 not too small.

215

216 In addition to the results presented in Table 2, we determined by simulation the
217 minimum detectable effect size f_{\min} – defined as the value of f for which the test at 0.05
218 significance level achieves a power of 0.8 – for the cases considered in Table 2. Results
219 are presented in Table 3. In overall terms, the rotation test has good power once f reaches
220 approximately 2.

221

222

223 **An Application to Sperm Whale Response to Airgun Sounds**

224

225 In this section, we apply the rotation test to some experimental data involving the
226 exposure of a sperm whale to air-gun sounds. Air-guns are a source of loud, impulsive
227 low-frequency underwater sound. They are generally deployed in towed arrays for
228 geophysical exploration (Richardson et al. 1995). Air-gun arrays have very high source
229 levels (Richardson et al. 1995; Caldwell & Dragoset 2000) and there is a concern that
230 exposing sperm whales and other marine mammals to air-gun noise may have adverse
231 impacts on their behavior (Gordon et al. 2003).

232

233 Because sperm whales use echolocation to locate prey, one hypothesized behavioral
234 impact of air-gun sound is a reduction in whale foraging rate. Sperm whales produce
235 regular echolocation clicks almost continuously while foraging, interrupted only by short
236 pauses and buzzes (short series of rapid echolocation clicks indicative of attempted prey
237 capture (Whitehead 2003, Miller et al. 2004a)). Whales begin producing echolocation
238 clicks during the descent phase of deep dives, stop clicking during or just prior to ascent,

239 and do not generally produce series of regular echolocation clicks while at the surface or
240 during shallow dives (Watwood et al. 2006). We therefore defined foraging periods as
241 the portions of deep dives between the start and end of regular echolocation clicks. The
242 behavioral event of interest was the production of echolocation buzzes, which serve as a
243 proxy for foraging rate.

244

245 The data used here were collected during controlled exposure experiments
246 conducted on the 2002 and 2003 Sperm Whale Seismic Study cruises. During the
247 experiments, dtags (Johnson & Tyack 2003) were attached to individual whales to record
248 sound and movement data during control conditions (no air-gun sound exposure) and
249 treatment conditions (air-gun sound exposure). Air-guns were fired every 15 seconds
250 during the treatment period. Detailed information on these experiments can be found in
251 Jochens & Biggs (2003, 2004) and Madsen et al. (2006). Here, we present data from a
252 single individual.

253

254 The behavioural record is shown in Figure 1. For this whale, the total time spent
255 foraging during the observation period was $T = 5.89$ hours, of which $T_C = 4.74$ hours was
256 under control conditions (the air-gun array was not operating) and $T_T = 1.15$ hours was
257 under treatment conditions (the air-gun array was in operation). A total of $n = 153$
258 echolocation buzzes were recorded, of which $N_C = 139$ occurred during control
259 conditions and $N_T = 14$ occurred during treatment conditions. The empirical rate of
260 events during control conditions was $29.3 \text{ events h}^{-1}$ and the corresponding rate during

261 treatment conditions was only 12.2 events h^{-1} . The value of the binomial test statistic in
262 (1) is 3.43, which is significant at approximately the 0.0006 level.

263

264 For reasons connected to the spatial distribution of prey and whale foraging
265 behavior, we expected that the Poisson model underlying the binomial test was unlikely
266 to apply to this time-series of sperm whale foraging events. This expectation was
267 confirmed by an analysis of the intervals between events, which revealed positive
268 autocorrelation at short lags. As the intervals in a Poisson process are independent, this
269 autocorrelation is evidence of non-Poisson behaviour in the point process. We therefore
270 applied the rotation test to these data. The histogram of values of N_C based on 10,000
271 rotations is shown in Figure 2. Of these, 647 exceeded the observed value of 139 for an
272 estimated two-sided significance level of approximately 0.13. In contrast to the binomial
273 test, by conventional standards, the null hypothesis cannot be rejected by the rotation test.
274 It is not possible to calculate a priori power estimates for the sperm whale data set, since
275 we do not know the true process generating the buzz time series data, and thus we can not
276 produce the simulated data needed for power estimation.

277

278 **Discussion**

279

280 The rotation test is a general nonparametric approach that can be used when data exhibit
281 serial dependence. The purpose of this paper has been to describe, evaluate, and illustrate
282 this test in the specific context of testing for a treatment effect on the rate of a
283 behavioural point process. We have shown that, in this context, the rotation test works

284 well, maintaining the nominal significance level while providing high power when the
285 data do not follow a Poisson process. In contrast, the binomial test is invalid in this case.

286

287 A common approach to analyzing behavioural point processes is to reduce the
288 data to empirical rates within time bins (e.g, Cherry 1989; Mooring 1995; Paredes et al.
289 2005; Fernández-Juricic & Tran 2007). As a general proposition, binning point process
290 data entails a loss of power (Dean & Balshaw 1997) and is not recommended. Moreover,
291 the analysis of binned data is also affected by non-Poisson behaviour in the underlying
292 point process. Briefly, if the underlying point process is Poisson, then the counts within
293 bins will have Poisson distributions. Statistical methods for analyzing Poisson count data
294 are reviewed in McCullagh & Nelder (1989). However, if the underlying point process is
295 not Poisson, then the distribution of bin counts is also not Poisson and the results of these
296 methods can be misleading (Paul & Banerjee 1998). A common alternative to the
297 Poisson distribution for count data is the negative binomial distribution. Parametric
298 methods for analyzing negative binomial data are available (e.g., Barnwal & Paul 1988;
299 Paul & Banerjee 1998). The rotation test provides a nonparametric alternative.

300

301 Turning to the results of the previous section, it is clear that no general conclusion
302 about the effect of air-gun noise on sperm whales can be drawn from the results of the
303 single test presented here. However, it is also worth pointing out that the hypothesized
304 effect of air-gun noise is a *reduction* in foraging. Had a one-sided test for such a
305 reduction been performed, the significance level would have been around 0.065 which, in
306 light of power considerations, is certainly suggestive of an effect.

307

308 Finally, although this paper has focused on the rotation test in the context of
309 analyzing behavioural point process data, the same general method could be used in other
310 situations. For example, Shapiro (2008) used a rotation test to determine whether the
311 frequencies of different types of vocalizations in killer whales differed between
312 behavioural states. In this case, the approach was used to account for serial dependence
313 in vocalization type.

314

315 **Acknowledgements**

316

317 We are grateful to Peter Tyack for advice and helpful comments. We thank the crews and
318 science parties of the research cruises on which sperm whale airgun exposure data were
319 collected, especially Peter Tyack and Patrick Miller. Funding for the sperm whale
320 experiments was provided by the Office of Naval Research, the U.S. Department of the
321 Interior Minerals Management Service Cooperative Agreements Nos. 1435-01-02-CA-
322 85186 and NA87RJ0445, and the Industry Research Funding Coalition. All approaches to
323 animals for tagging were made following the conditions of NMFS research permits 981-
324 1578 and 981-1707. The Woods Hole Oceanographic Institution Animal Care and Use
325 Committee approved this research.

326

327 **References**

- 328 **Barnwal, R.K. & Paul, S.R.** 1988. Analysis of one-way layout of count data with
329 negative binomial variation. *Biometrika*, 75, 215-222.
- 330 **Caldwell, J. & Dragoset, W.** 2000. A brief overview of seismic air-gun arrays. *The*
331 *Leading Edge*, 19, 898-902.
- 332 **Cherry, J. A.** 1989. Ultrasonic Vocalizations by Male Hamsters - Parameters of Calling
333 and Effects of Playbacks on Female Behavior. *Animal Behaviour*, 38, 138-153.
- 334 **Cox, D. R. & Lewis, P. A. W.** 1978. *The statistical analysis of series of events*, 2nd edn.
335 London: Chapman & Hall.
- 336 **Dean, C.B. & Balshaw, R.** 1997. Efficiency lost in analyzing counts rather than event
337 times in Poisson and over-dispersed Poisson regression models. *Journal of the American*
338 *Statistical Association*, 92, 1387-1398.
- 339 **Fernández-Juricic, E. & Tran, E.** 2007. Changes in vigilance and foraging behaviour
340 with light intensity and their effects on food intake and predator detection in house
341 finches. *Animal Behaviour*, 74, 1381-1390.
- 342 **Gordon, J., Gillespie, D., Potter, J., Frantzis, A., Simmonds, M. P., Swift, R. &**
343 **Thompson, D.** 2003. A review of the effects of seismic surveys on marine mammals.
344 *Marine Technology Society Journal*, 37, 16-34.
- 345 **Haccou, P. & Meelis, E.** 1992. *Statistical Analysis of Behavioural Data*. New York:
346 Oxford University Press.

- 347 **Harkness, R. D. & Isham, V.** 1983. A bivariate spatial point pattern of ants' nests.
348 *Applied Statistics*, 32, 293-303.
- 349 **Jochens, A. E. & Biggs, D. C.** 2003. Sperm whale seismic study in the Gulf of Mexico:
350 Annual report, year 1. New Orleans, LA: U.S. Department of the Interior Minerals
351 management service, Gulf of Mexico OCS Region.
- 352 **Jochens, A. E. & Biggs, D. C.** 2004. Sperm whale seismic study in the Gulf of Mexico:
353 Annual report, year 2. New Orleans, LA: U.S. Department of the Interior Minerals
354 management service, Gulf of Mexico OCS Region.
- 355 **Johnson, M. P. & Tyack, P. L.** 2003. A digital acoustic recording tag for measuring the
356 response of wild marine animals to sound. *IEEE Journal of Oceanic Engineering*, 28, 3-
357 12.
- 358 **Lawrance, A. J. & Lewis, P. A. W.** 1979. Simulation of some autoregressive Markovian
359 sequences of positive random variables. In: *Proceedings of the IEEE 11th Conference on*
360 *Winter Simulation*, pp. 301-307. San Diego, CA, USA: IEEE.
- 361 **Madsen, P. T., Johnson, M., Miller, P. J. O., Soto, N. A., Lynch, J. & Tyack, P.** 2006.
362 Quantitative measures of air-gun pulses recorded on sperm whales (*Physeter*
363 *macrocephalus*) using acoustic tags during controlled exposure experiments. *Journal of*
364 *the Acoustical Society of America*, 120, 2366-2379.
- 365 **McCullagh, P. and J. A. Nelder.** 1989. *Generalized Linear Models*. New York:
366 Chapman & Hall.
- 367 **Miller, P. J. O., Johnson, M. P. & Tyack, P. L.** 2004a. Sperm whale behaviour

368 indicates the use of echolocation click buzzes 'creaks' in prey capture. *Proceedings of the*
369 *Royal Society of London Series B-Biological Sciences*, 271, 2239-2247.

370 **Miller, P.J.O., Shapiro, A.D., Tyack, P.L. & Solow, A.R.** 2004b. Call-type matching
371 in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Animal*
372 *Behaviour*, 67, 1099-1107.

373 **Mooring, M. S.** 1995. The effect of tick challenge on grooming rate by impala. *Animal*
374 *Behaviour*, 50, 377-392.

375 **Paredes, R., Jones, I. L. & Boness, D. J.** 2005. Reduced parental care, compensatory
376 behaviour and reproductive costs of thick-billed murrelets equipped with data loggers.
377 *Animal Behaviour*, 69, 197-208.

378 **Paul, S.R. and Banerjee, T.** 1998. Analysis of two-way layout of count data involving
379 multiple counts in each cell. *Journal of the American Statistical Association*, 93, 1419-
380 1429.

381 **Richardson, W. J., Greene, C. R., Jr., Malme, C. I. & Thompson, D. H.** 1995. *Marine*
382 *Mammals and Noise*. San Diego, CA: Academic Press.

383 **Shapiro, A.D.** 2008. Orchestration: The movement and vocal behavior of free-ranging
384 Norwegian killer whales (*Orcinus orca*). PhD thesis, MIT/WHOI.

385 **Sibly, R.M., Nott, H.M.R. & Fletcher, D.J.** 1990. Splitting behaviour into bouts.
386 *Animal Behaviour*, 39, 63-69.

387 **Slater, P.J. & Lester, N.P.** 1982. Minimising errors in splitting behavior into bouts.

388 *Behaviour*, 79, 153-161.

389 **Stephens, M. A.** 1986. Tests for the uniform distribution. In: *Goodness-of-fit techniques*
390 (Ed. by D'Agostino, R. & Stephens, M. A.), pp. 331-365. Boca Raton, FL: CRC Press.

391 **Thomas, M.** 1949. A generalization of Poisson's binomial limit for use in ecology.
392 *Biometrika*, 36, 18-25.

393 **Watwood, S. L., Miller, P.J.O., Johnson, M., Madsen, P.T. & Tyack, P.L.** 2006.
394 Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *Journal of*
395 *Animal Ecology*, 75, 814–825.

396 **Whitehead, H.** 2003. *Sperm Whales: Social Evolution in the Ocean*. Chicago, IL:
397 The University of Chicago Press.

398

399 **Tables**

400

401 Table 1. Validity of the binomial and rotation tests

402

403

overall mean rate

404

500

1000

405

Poisson

406

binomial

0.05

0.05

407

rotation

0.042

0.054

408

409

Thomas

410

binomial

0.221

0.233

411

rotation

0.045

0.049

412

413

EAR

414

binomial

0.247

0.251

415

rotation

0.052

0.052

416

417 The rate at which the null hypothesis of no treatment effect was falsely rejected in testing

418 at the 0.05 significance level using the binomial test and the rotation test for data

419 simulated from the Poisson, Thomas, and EAR models with overall mean rates of 500

420 and 1000. For the Thomas model, $\theta = 1$ and for the EAR model $\rho = 0.5$. Results are

421 based on 1000 simulations except for the binomial test under the Poisson model where
422 the theoretical result is given.

423

424 Table 2. Power of the rotation test

425

426

overall mean rate

427

500

1000

428

f

f

429

1.5

2

3

1.5

2

3

430

Poisson

431

binomial

1

1

1

1

1

1

432

rotation

0.61

0.90

0.99

0.82

0.98

1

433

434

Thomas

435

rotation

0.34

0.63

0.88

0.56

0.86

0.97

436

437

EAR

438

rotation

0.37

0.64

0.86

0.54

0.83

0.98

439

440

The power of the rotation test at the 0.05 significance level is presented, for data

441

simulated under the Poisson, Thomas, and EAR models when the mean rate under

442

treatment is a factor *f* greater than that under control and when the overall mean rate is

443

fixed at 500 and 1000. For the Thomas model, we always used $\theta = 1$; for the EAR model

444

we always used $\rho = 0.5$. For the Poisson model, results are also given for the binomial

445

test. Results are based on 1000 simulations.

446

447 Table 3. Minimum detectable effect levels.

		overall mean rate	
		500	1000
450	Poisson		
451	binomial	1.3	1.2
452	rotation	1.7	1.5
453			
454	Thomas		
455	rotation	2.5	1.8
456			
457	EAR		
458	rotation	2.7	2.0
459			

460 The minimum detectable effect size f_{\min} – defined as the value of f for which the test at
461 0.05 significance level achieves a power of 0.8 – for the cases considered in Table 2.
462 Results are based on 1000 simulations.

463 **Figure Legends**

464

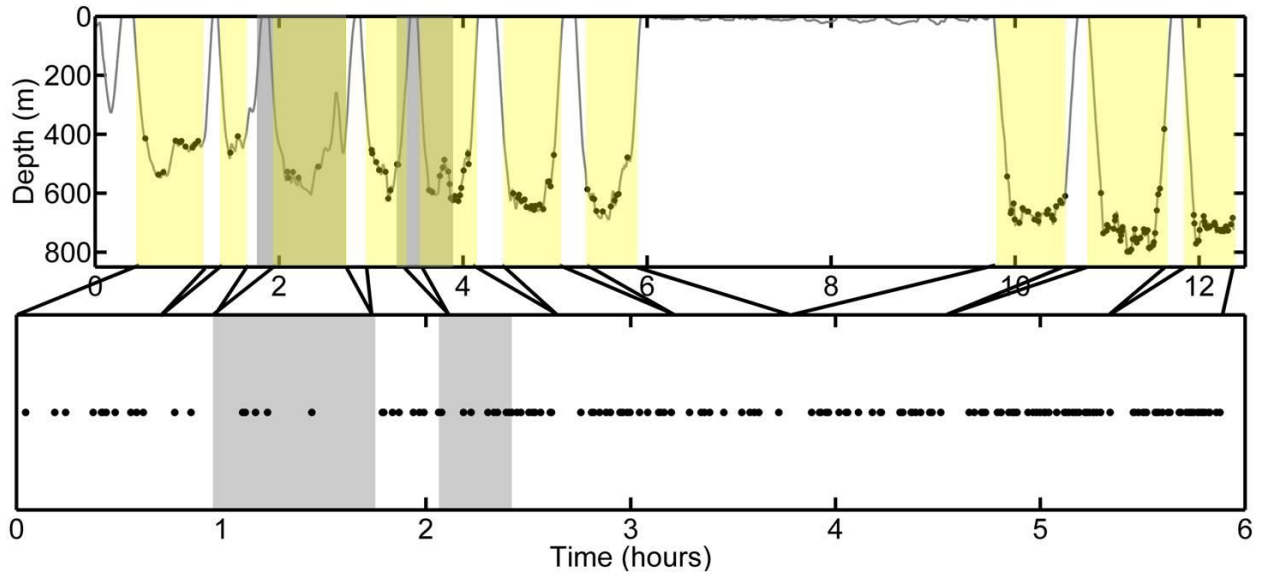
465 Figure 1. Top panel: Dive profile of the tagged sperm whale. The grey line indicates
466 whale depth, and black circles indicate the times of echolocation buzzes. Airgun
467 exposure periods are shaded gray. Black lines connecting the top and bottom panels
468 illustrate how dive ascents, descents and surface periods were cut from the dataset to
469 produce the buzz time-series we analyzed. Only foraging periods (indicated by yellow
470 shading) were included in the buzz time-series. Bottom Panel: Time-series of
471 echolocation buzzes produced by the sperm whale during foraging periods. Black dots
472 indicate the times of buzzes, and gray shaded areas indicate airgun exposure periods.

473

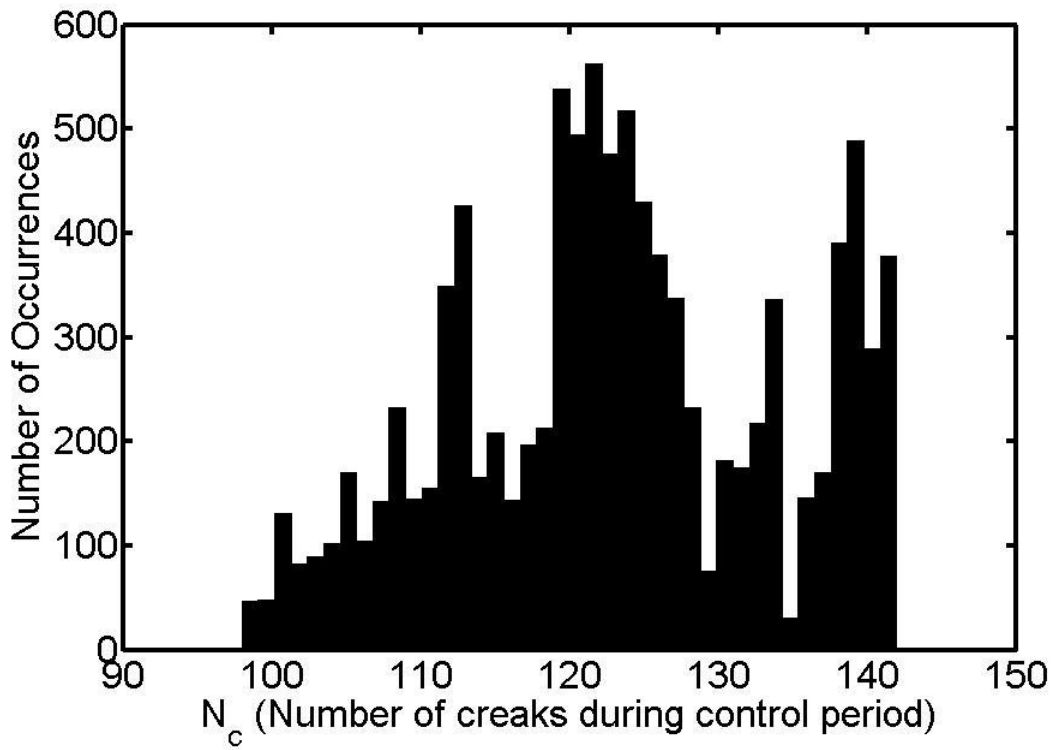
474 Figure 2. Histogram of values for N_C , the number of creaks during the control period,
475 obtained in 10,000 rotations of the sperm whale dataset. (The value of N_C was 139 for
476 the original, non-rotated dataset.)

477

478 **Figures**



479



480

481

482