



This is a peer-reviewed, post-print (final draft post-refereeing) version of the following published document and is licensed under Creative Commons: Attribution-Noncommercial-No Derivative Works 3.0 license:

Stafford, Richard and Hart, Adam G and Goodenough, Anne E (2013) A visual method to identify significant latitudinal changes in species' distributions. *Ecological Informatics*, 15. pp. 74-84. ISSN 15749541

Official URL: <http://dx.doi.org/10.1016/j.ecoinf.2013.03.003>

DOI: <http://dx.doi.org/10.1016/j.ecoinf.2013.03.003>

EPrint URI: <http://eprints.glos.ac.uk/id/eprint/3362>

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.

This is a peer-reviewed, post-print (final draft post-refereeing) version of the following published document:

**Stafford, Richard and Hart, Adam G
and Goodenough, Anne E (2013). *A visual method to
identify significant latitudinal changes in species'
distributions*. Ecological Informatics, 15, 74-84. ISSN
15749541**

Published in Ecological Informatics, and available online at:

<http://www.sciencedirect.com/science/article/pii/S1574954113000241?np=y>

We recommend you cite the published (post-print) version.

The URL for the published version is <http://dx.doi.org/10.1016/j.ecoinf.2013.03.003>

Disclaimer

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT

1 **Inferential and visual analysis of ethogram data using multivariate techniques**

2 Richard Stafford^{a*}, Anne E. Goodenough^b, Kathy Slater^c, William Carpenter^b, Laura Collins^b, Heather
3 Cruickshank^b, Sarah Downing^b, Sally Hall^b, Katie McDonald^d, Heather McDonnell^b, Bryony Overs^b,
4 Lizzie Spiers^b, Mark Stetter^e, Heike Zitzer^c, Adam G. Hart^b

5

6 a) Luton Institute of Research in Applied Natural Sciences, University of Bedfordshire

7 b) Department of Natural and Social Sciences, University of Gloucestershire, Cheltenham

8 c) Operation Wallacea, Spilsby, Lincolnshire

9 d) West Midlands Safari Park, Bewdley, Worcestershire

10 e) Animals Science and Environment, Walt Disney World, Lake Buena Vista, Florida

11

12 * Corresponding author. Email: rick.stafford7@gmail.com

13 Division of Science, University of Bedfordshire, Luton, LU1 3JU, UK.

14 Tel: +44 (0) 1582 489158

15 Fax: +44 (0) 1582 489212

16 Word count =

17

18

19 **Inferential and visual analysis of ethogram data using multivariate techniques**

20 Activity budgets are frequently used to examine behaviours of animals, especially of large mammals
21 in field or captivity conditions (e.g. Altmann 1974; Weller and Bennett 2001; MacNulty et al. 2007).
22 Often, such processes are conducted using ethograms, where a number of typical behaviours are
23 listed (such as foraging, sleeping, walking, standing still, interacting with others) and either the
24 duration of each behaviour within each observation period is noted, or, more normally, the
25 occurrence of a certain behaviour is recorded at a regular time interval (Dawkins 2007; Martin and
26 Bateson 2007). The technique is simple, and clearly effective in calculating the proportion of time
27 spent undertaking each of the behaviours. However, analysis of the data is problematic (Ramson and
28 Cade 2009). Even if the same animal is repeatedly sampled (for example on different days), the
29 averages and some measure of variability or precision are normally calculated for each of the
30 behavioural categories included in the ethogram separately (Ramson and Cade 2009). While
31 inferential statistics could be used to calculate significant differences between individuals in terms of
32 the occurrence of a specific behaviour, there are problems with the independence of these data
33 both in terms of repeated measures, and because all behaviours must sum to 1 as they are mutually
34 exclusive— see Aitchison 1986 and Underwood 1996 for more details about these points). Even if
35 such strict limitations on data analysis are relaxed, then this still only indicates whether animal X
36 conducts behaviour A more or less frequently than animal Y.

37 Because of these issues, it would be preferable to use a multivariate method to analyse the
38 overall behaviour of individuals (defined as all behaviours in the activity budget e.g. Mielke and
39 Berry 2007) and to compare it to other individuals for whom identical data are held. Principal
40 Component Analysis (PCA), and associated plotting of resultant components in 2 or 3 dimensions, is
41 one possible method (i.e. biplots, where any given case is plotted against the first two principal
42 components). This can give an indication of how different animals behave, based on all the
43 behaviours examined. However, several limitations to this technique exist. It is generally

44 recommended that the case to variable ratio for PCA is > 3:1, that is the number of observed animals
45 should be >3 times the number of behaviours examined (Tabachnick and Fidell, 1989), and that
46 ideally the number of cases should be high (> 300, Comrey and Lee 1992). Given that most
47 ethograms include a large number of different behaviours, and the number of animals studied is
48 often small, these limitations are significant. It is possible to use replicate sampling of the same
49 animal to boost the number of cases (using each replicate sampling period as a separate case), but
50 differences are likely to occur in behaviours based on factors such as time since eating, proximity of
51 other individuals of the same or opposite sex, hormonal changes or seasonal changes. Furthermore,
52 with traditional PCA techniques, it is not possible to determine whether differences in behaviour are
53 statistically significant or not (despite techniques such as concentration ellipses, which do not give a
54 good indication of statistical differences). Theoretically, if a biplot indicates clustering of cases from
55 one animal, and distinct, separate clustering of cases from a second animal, then they are likely to
56 be different, but, in practice, points are often interspersed and overlap with one another for the
57 reasons mentioned previously. As such, judging differences in behaviour becomes very subjective
58 (Gabriel 1971).

59 A method of combining inferential statistics with PCA has recently been developed, based on
60 constructing bootstrapped confidence intervals (or confidence radii since precision is calculated in
61 three dimensions) for each case in the PCA (Catlin-Groves et al. 2009). Because this technique
62 calculates the precision of the mean using confidence intervals, many limitations of PCA, such as the
63 case to variable ratio are less important, since lack of precision on the PCA axes is indicated by
64 increased confidence intervals. Furthermore significant differences can be inferred on the basis of
65 whether confidence radii overlap (Catlin-Groves et al. 2009). As such, the technique should be
66 beneficial for application to activity budget behavioural data collected through ethogram studies.

67 Here we develop the framework for applying this technique to activity budget studies, and
68 show the results of its application to four studies (captive and non-captive mammals, and
69 invertebrates) that indicate its potential broad application.

70 **Field data collection**

71 *Tigers in captivity*

72 Data collection took place at West Midland Safari Park in Bewdley, Worcestershire, UK (52°22'51" N,
73 2°17'06" W). In total, four Bengal tigers (*Panthera tigris tigris*) were studied, in two pairs. Each pair
74 cohabited permanently, and was moved around a number of enclosures on a day-by-day basis. The
75 first pair (tigers 1 and 2) was an unrelated male-female pair and the second pair (3 and 4) was a male
76 – female sibling pair. The enclosures in which the tigers were studied contained trees and a dual
77 layered platform in the centre of the compound. One of the enclosures also contained a small pool.

78 Data were collected in 1 h or 2 h periods, with behaviours recorded on an ethogram (Table
79 1) at 30 second intervals. In total 12 h of data were collected for each tiger (with a data point
80 collected from pairs of tigers simultaneously), giving 1440 ethogram observations per tiger.

81

82 *Elephants in a nature reserve*

83 This study was conducted at the 73.6km² Pongola Nature Reserve in South Africa (27°28'18"S,
84 31°56'49"E). Data were collected on five adult males using instantaneous scan sampling at 5 min
85 intervals (as per Altmann 1974). At each scan the behaviour of each elephant was recorded using the
86 behavioural categories listed in Table 2. Data from each male was collected until the male's
87 behaviour could no longer be accurately visually identified using binoculars. In total 154 data
88 collection points were collected for the five elephants, with a minimum of 22 ethogram observations
89 per individual.

90

91 *Dogs in rescue shelters*

92 Dogs were studied at Cheltenham Animal Shelter, Gloucestershire, UK (51° 54' 50.84" N, 2° 4' 59.51"
93 W). Dogs had already been assigned a traffic light coding of behaviour with red dogs being
94 aggressive and green dogs being friendlier and with fewer behavioural problems. This coding was
95 decided from a preliminary behaviour assessment by the shelter staff when the dogs entered the
96 shelter. Three red dogs and three green dogs were observed while being exercised in the shelter's
97 paddock. Each dog was observed three times for a total of 20 minutes, and behaviours noted every
98 20s from the list in Table 3. In total 180 ethogram observations were collected from each dog.

99

100 *Shore crab behaviour to a simulated predator*

101 Crabs were collected from a rockpool at Crantock Beach in Cornwall, UK (50°24'20" N, 5°07'51" W).
102 The rockpool was ~ 2.5 m above chart datum. For each trial, three crabs were transferred to a 1 m
103 diameter plastic experimental arena (filled with 10 cm depth of freshly collected seawater), located
104 *in situ* next to the rock pool, and allowed to acclimatise for 1 h before being observed for 10 mins.
105 During this 10 min period, crab behaviour was recorded every 30 s from the list of behaviours in
106 Table 4. Crabs were placed in groups of either three adult crabs (carapace width > 40 mm) or
107 juvenile crabs (carapace width > 20 but < 40 mm) and for each group, they were either left free from
108 visual disturbance over the 10 min period or were presented with a shadow of a predator (a
109 silhouette of a seagull) for 10 s at 60 s intervals. In total 24 crabs were used, hence each of the four
110 treatments (adult or juvenile, in the presence or absence of a visual predator stimulus) was
111 replicated twice. Each crab had 20 ethogram observations. After the study, crabs were released back
112 into the rockpool from which they came. In no cases were crabs removed from their natural
113 environment for more than 2 h.

114 **Statistical methods**

115 The bootstrapped PCA process was derived from that described in Catlin-Groves et al. (2009) and
116 slightly modified here for use on behavioural datasets. The code runs in the R statistics environment
117 (R Core Development Team 2011) and is available as supplementary material to this paper, along
118 with a sample dataset used in this study (the tiger dataset).

119 For each analysis, a frequency distribution table was set up for all cases using a spreadsheet.
120 A unique classifying number for each behaviour in the ethogram was assigned (e.g. for Table 1, '1'
121 would be assigned to feeding, '2' to foraging and so on). This classification number was typed into
122 the spreadsheet in a vertical column, with the number of entries corresponding to the percentage
123 frequency of that behaviour. For example, if behaviour 1 occurred 32% of the time, the term '1'
124 appeared in the first 32 rows of the spreadsheet. As such, each case is inputted in separate columns,
125 and behaviours indicated in rows 1-100. The number of different columns was equal to the number
126 of cases being considered within a specific analysis. The term 'case' is defined by the user. In most
127 studies here, it is the combined ethograms from any individual animal, over all the sampling periods,
128 but could be combined data from ethograms for an individual on days it had been fed, as compared
129 to days it had not been fed, for example, or multiple individuals within a particular category such as
130 sex. This classification of 'case' is considered in greater detail in the discussion and examples of
131 different classifications of case are given in the results. The conversion of behaviour into
132 percentages is to ensure that there were always 100 data points in each sample, and allow
133 consistent rules to be formulated (such as the size of the subsample for bootstrapping) to apply the
134 technique generally to behaviours where the number of observations can vary (as per the studies
135 considered here).

136 From each case, 100 points were randomly taken, with replacement, to obtain a sample of
137 the behaviour (the use of 100 points – with replacement – from 100 does not imply all points are
138 sampled each time, and is the overwhelming consensus of sample size for bootstrapping in the

139 literature – e.g. Efron 1979; Crawley 2007; Martínez-Muoz and Suárez 2010). Using the ‘prcomp’
 140 function in R, the first three principal components of each sample were calculated and stored for
 141 each case, and the process repeated 10 000 times. A mean value of the 10 000 replicates was
 142 calculated and 95% confidence limits were calculated by excluding the highest and lowest 2.5 % of
 143 the values (Crawley, 2005). By altering this parameter to the highest and lowest 5% or 10%,
 144 confidence limits can be obtained at 90% or 80% levels, respectively. Upper and lower confidence
 145 intervals for all three of the stored principal components were averaged to give a confidence radius.
 146 The mean values of the principal components for each site were plotted in 3 dimensions and the
 147 confidence radius indicated the size of the sphere, or bubble. Plots were made using the RGL library
 148 and rgl.sphere function for R (Adler and Murdoch 2008). However, because of some issues of how
 149 principal components are calculated, the following modifications were required to produce the
 150 bootstrapped means and confidence radii.

151 Initially, the full dataset was analysed using the ‘prcomp’ function to give a baseline value for
 152 each case. For each replicate run of the bootstrapped principal components (where n = 100; but
 153 sampled with replacement), the full dataset (where n = 100; but without replacement) for each case
 154 was also analysed, essentially doubling the cases in replicate run. By calculating a vector to
 155 transform each point from the full dataset back to its corresponding baseline point (equation 1), and
 156 then applying the same vector to the bootstrap points (equation 2), the variability in the
 157 bootstrapped points is restricted to variation between differences in the placement of points on the
 158 initial principal component axes, and not variation between both the placement of points and
 159 alignment of principal component axes. So:

$$160 \quad v_{[x,y,z]} = I_{[x,y,z]} - \hat{I}_{[x,y,z]} \quad [1]$$

$$161 \quad B_{mod[x,y,z]} = B_{calc[x,y,z]} + v_{[x,y,z]} \quad [2]$$

162 where v is the vector, l is the initial full data point calculated without the addition of the bootstrap
163 points, i is the full data point calculated along with the bootstrap points, B_{mod} is the bootstrapped
164 point modified by the vector and B_{calc} is the bootstrap point calculated directly by PCA.

165 Applying this vector also accounted for the arbitrary sign applied to the magnitude of the principal
166 component (during replicates on identical datasets, the value of a point on a principal component
167 axis could be assigned as 1 or -1). The vector transformation eliminated this problem unless the sign
168 (+ or -) of the full dataset differed from the sign of the bootstrapped dataset for the same point. If
169 this was the case, the magnitude of the vector in this dimension was ~ 2 x that of the magnitude of
170 the value of the full dataset point. To account for this problem, if the magnitude of the vector
171 exceeded 1.2 x that of the magnitude of the value of the full dataset point, the magnitude of the
172 vector in this dimension was calculated by adding the two points (equation 3) and then subtracting
173 the calculated bootstrap value from the vector (equation 4).

174

$$175 \quad v_{[x,y,z]} = l_{[x,y,z]} + i_{[x,y,z]} \quad [3]$$

$$176 \quad B_{mod[x,y,z]} = v_{[x,y,z]} - B_{calc[x,y,z]} \quad [4]$$

177

178 The value of 1.2 x the magnitude as the demarcation between equations 1 and 3 being applied was
179 previously been shown to be suitable, and sensitivity analysis of the results indicate that values
180 between 1 and 1.5 do not cause changes in output (Catlin-Groves et al. 2009).

181

182 **Results and Discussion**

183 *Tigers in captivity*

184 Using the standard 'prcomp' function on the full data set, the first three principal components were
185 shown to explain 99.0% of the total variance of the data set. Analysis of the four tigers showed that
186 the two females (2 and 4) had overlapping bubbles indicating that their behaviours were not
187 significantly different from each other. The two males had bubbles which also overlapped, but tiger
188 3 had a significantly different behaviour from tiger 4, but not from tiger 2 (Figure 1). Tiger 1 showed
189 significantly different behaviour compared to both females. The male tiger 3 was more similar in
190 behaviour to the two females than the male tiger 1 – which spent a considerable less time pacing
191 than the other three individuals. Indeed, Tiger 1 was recorded pacing on average 2.75 times per day,
192 compared to an average of 65.5 times per day with the other male, Tiger 3. While little work has
193 been conducted on sex specific behaviours in captive carnivores, some studies (e.g. Renner and
194 Lussier 2002) have found sex specific differences to certain aspects of captive carnivore behaviours.
195 The results from this study provide some support for sex specific differences in captive tiger
196 behaviour, but also indicate that variability between individuals may be as important as sex based
197 differences.

198 *Elephants in a nature reserve*

199 Using the standard 'prcomp' function on the full data set, the first three principal components were
200 shown to explain 92.1% of the total variance of the dataset. Analysis of the five bull elephants'
201 activity budgets using the bootstrapping methods showed no significant differences between
202 elephants at the 95 % confidence level, as there is overlap between all of the coloured bubbles that
203 represented the individual elephants (Figure 2a). Such a lack of difference in activity budgets may be
204 unsurprising, given that activity levels in the African savannah are heavily constrained by time spent
205 resting as a means of coping with heat stress (Dunbar 1992). Moreover, elephant activity at Pongola
206 is further constrained by limited food available to this population, which far exceeds the park's
207 carrying capacity. However, the distribution of the bubbles does correspond closely to the previously
208 determined dominance hierarchy of these bull elephants (H. Zitzer unpublished data), with the left

209 most elephant being the most dominant, and the dominance hierarchy decreasing from left to right
210 (Figure 2a). Given that dominance was calculated by aggressive interactions, and these data
211 presented in this study are from activity budgets (where dominance interactions are largely absent),
212 such a correlation of results is a good indication that the technique is incorporating many aspects of
213 the elephants' behaviour.

214 The plot of all five bull elephants can make determining significant differences between non-
215 adjacent individuals difficult. However, pairwise comparisons can also be plotted, without the
216 analysis being rerun. To minimise type I errors of pairwise comparisons, it is logical to examine the
217 furthest apart individuals first (here elephants 1 and 5), as per the procedure in standard post-hoc
218 tests such as Student-Newman-Keuls (SNK) tests. In this case, while no significant differences occur
219 at the 95% confidence level (Figure 2b), differences do occur at the 90% confidence level between
220 the overall activity budget of elephants 1 and 5 (Figure 2c). From an examination of the activity
221 budget data, it can be seen that the key differences in behaviour are an increase in resting and
222 feeding, and a decrease in moving in the most dominant elephant, as compared to the least
223 dominant (Elephant 1 – movement = 43%, resting = 26%, feeding = 25%; Elephant 5– movement =
224 58%, resting = 17%, feeding = 15%). The differences in activity budget between the highest and
225 lowest ranking male are in line with previous field observations of these elephants. The dominant
226 male spent nearly all of his time travelling with the larger of the two female herds. As he constantly
227 had access to females, the dominant male travelled less and spent more time resting and feeding
228 with the females. The subordinate male spent a significant amount of time alone wandering
229 between the two female herds attempting to gain access to the females and as a result spent
230 significantly more time moving than the dominant male (K. Slater and H. Zitzer unpublished data).

231

232 *Dogs in rescue shelters*

233 Using the standard 'prcomp' function on the full data set, the first three principal components were
234 shown to explain 92.1% of the total variance of the data set. The bubble plot displayed some
235 significant differences between dogs (Figure 3a). The clustered group of three bubbles represent the
236 red dogs, and the three separated bubbles represent the green dogs. There is a clear significant
237 difference between all three green dogs in relation to one another and each of the red dogs,
238 indicating that their initial behavioural classification could also be determined by activity budget
239 ethograms. By redefining the classes used here, it was also possible to determine if differences occur
240 between the red and green dogs studied in general. By combining all data on the three red and three
241 green dogs, the process can be rerun. This is case, pooling the data in this way demonstrates that
242 there is not an overall significant difference between the red and green dogs, despite each individual
243 green dog being different from all red dogs (Figure 3b), although again, a significant difference
244 occurs at 90% confidence (Figure 3c). As with the tiger data, such a response indicates that variability
245 between dogs (especially the green classified dogs) can be large. In this case, differences in green
246 dog behaviour are larger than between red dogs. This may be explained by the fact that red dogs are
247 classified by aggressive characteristics – hence all behave in an aggressive manner, whereas green
248 dogs display a more natural, and varied range of domestic dog behaviours.

249

250 *Shore crab behaviour to a simulated predator*

251 Using the standard 'prcomp' function on the full data set, the first three principal components were
252 shown to explain 99.7% of the total variance of the data set. Significant differences in behaviours
253 between the treatment groups were found at the 95% confidence level (Figure 4). Juvenile crabs
254 behaved in a similar way in the absence of a predator stimulus to adult crabs in the presence of the
255 predator stimulus. Both juveniles and adults showed a similar response to predators (a downwards
256 movement in the plot of behaviour in Figure 4). From a re-examination of the data, this tends to
257 indicate an increase in hiding behaviour from both juveniles and adults in the presence of a predator

258 (from 13 to 37 % of the time in mature crabs and from 47 to 75 % of the time in juveniles).
259 Differences in behaviour of crustaceans, especially in regard to life- and moult-cycle stage, are well
260 classified, with reduced locomotion and feeding activity at the most vulnerable stages (e.g. Lipcius
261 and Herrnkind 1992), hence while both adult and juvenile respond to a predator stimulus by hiding,
262 they start from different baseline activity behaviours.

263

264 *The statistical methods*

265 The technique of bootstrapping PCA analysis works well on the examples of activity budget /
266 ethogram-recorded behaviours studied here. The technique is flexible as regards: the number of
267 samples taken per animal, the confidence level examined, and, to a large extent, the definition of
268 'case', which could be an individual animal, or a group of animals (of the same sex, age group or any
269 other logical classification). However, there are some potential considerations and
270 recommendations for the application of the technique.

271 Firstly, the number of ethogram recordings used (or the sample size) must be large enough
272 to provide a good estimate of the activity budget of the animal studied. While the conversion of
273 different behaviours to percentages (hence the effective sample size is always 100) will not affect
274 the confidence interval size of a bootstrap method, clearly, limited recording may not capture the
275 full behaviour of the animal, as such, it is best to use similar sample sizes for different animals in the
276 study and to report the sample sizes used in the methods or results.

277 Secondly, all the data sets considered here had very large proportions of variability
278 explained by the first three principal components (> 90% in all cases). This means that the positions
279 of the bubbles on three dimensional plots are accurate simplifications of the multivariate complexity
280 inherent in the original data. If the proportion of variability explained by the first three principal
281 components decreases, the number of dimensions required of the plots needs, theoretically, to

282 increase – although this would make visual interpretation of the data very difficult. As such it is
283 recommended that this technique only be used where > 90% of the variability in the data is
284 explained by the first three principal components (this figure also follows standard practice
285 recommendations for biplots given in Crawley 2007).

286 Thirdly, the technique will naturally face some of the disadvantages of all confidence interval
287 methods as compared to inferential statistical hypothesis tests (Lanzante 2005). For example,
288 confidence interval estimation for univariate methods is not as powerful as equivalent t-tests or
289 ANOVA, at least when the data fulfil parametric assumptions. However, following the procedures
290 derived for ANOVA post-hoc tests, which involve testing the most different cases first, reduces the
291 number of pairwise comparisons which need to be made (see elephant example above).
292 Furthermore, corrections to eliminate type I error could easily be made by increasing the level of
293 significance from 95%, as per Bonferroni corrections or that occur in the standard Tukey test,
294 although this should be undertaken with caution since many authors advise against such
295 modifications due to the unproportional risk of type II error over minimising type I error (e.g.
296 Underwood 1996). Whether or not such changes to confidence limits need to be made depends on
297 the study in question, and whether interpretation of results is most sensitive to falsely detecting
298 differences, or not detecting real differences. While these modifications can help prevent issues of
299 type I error, the problems of pooling estimates of variability to a common standard deviation, which
300 can result in type 2 errors (the type most frequently found with the use of confidence interval
301 analysis - Lanzante 2005) do not apply to bootstrapping processes, where confidence intervals are
302 estimated directly from the data, and do not require an estimation of standard deviation.
303 Furthermore, the bootstrapping process does not necessarily result in symmetrical confidence
304 intervals around the mean, making the technique robust to the assumptions for parametric statistics
305 such as normally distributed data. Therefore, in many ways, the bootstrapping method detailed here
306 is more robust than many statistics for hypothesis testing, which require the homogeneous standard
307 deviations and normally distributed data between cases (Underwood 1996).

308 Finally, the issue of selecting a ‘case’ is not as advanced as for some statistical techniques. In
309 normal PCA, a case would correspond to a single observation period. Here, multiple observation
310 periods of a single individual can be combined as a case, as can multiple observation periods of many
311 individuals within a group (providing the replication of directly observed behaviour proposed by
312 Dawkins 2007). While this provides a flexible framework for hypothesis testing, a parallel can be
313 drawn with nested designs in general linear models. Nesting hierarchical responses (i.e.
314 observations of the same individual are nested within each individual, individuals of the same sex
315 are then nested within sex), rather than simply combining responses across all levels would,
316 potentially, allow differences in individuals, as well as differences between higher level ‘cases’ to be
317 determined in a single analysis, and allow an understanding of where the greatest variability lies (i.e.
318 between a behavioural category, between individuals or between replicate measures of the same
319 individual). However, such an approach would not present data in such a visually simple manner,
320 and in some cases, nesting factors within others produces less powerful inferential tests than not
321 conducting this nesting process (Hernández-Sánchez et al. 2003). A method of including nesting
322 would be a useful future improvement to this technique, however, it would also create an additional
323 level of complexity in performing the analysis, which in most cases, would not make a significant
324 difference to the outcome of the analysis.

325 The technique presented here provides an excellent framework for visualising activity
326 budget collected data and provides a novel method for determining significant differences between
327 classifications of interest within the dataset. While there are some residual issues in the application
328 of the technique, which necessitate researchers to think through analysis and interpretation of
329 resultant plots carefully, the method is a vast improvement on the statistical methods currently used
330 for such analysis.

331

332 **Acknowledgements**

333 The authors would like to thank Heinz Kohrs, director of ‘Space for
334 Elephants’ project and Cheltenham Animal Shelter for their assistance in with this study.

335

336 **References**

337 **Adler, D. & Murdoch, D.** 2008. *rgl: 3D Visualization Device System (OpenGL)*. R Package Version 0.77.
338 <http://rgl.neoscientists.org>.

339 **Aitchison, J.** 1986. The statistical analysis of compositional data (with discussion). *Journal of the*
340 *Royal Statistical Society (B)*, **44**, 1313–1325.

341 **Altmann, J.** 1974. Observational study of behavior: sampling methods. *Behaviour*, **49**, 227-267.

342 **Catlin-Groves, C.L., Kirkhope, C.L., Goodenough, A.E. & Stafford, R.** 2009. Use of confidence radii to
343 visualise significant differences in principal components analysis: application to mammal
344 assemblages at locations with different disturbance levels. *Ecological Informatics*, **4**, 147–151.

345 **Comrey, A.L. & Lee, H.B.** 1992. *A First Course in Factor Analysis*. 2nd Edition. Hillsdale, NJ: Lawrence
346 Erlbalm Associates.

347 **Crawley, M.J.** 2005. *Statistics: an Introduction using R*. Chichester, UK: Wiley.

348 **Crawley, M.J.** 2007. *The R Book*. Chichester, UK: Wiley.

349 **Dawkins, M.S.** 2007. *Observing animal behaviour: design and analysis of quantitative data*. Oxford:
350 Oxford University Press

351 **Dunbar, R.I.M.** 1992. Time: a hidden constraint on the behavioural ecology of baboons. *Behavioural*
352 *Ecology and Sociobiology*, **31**, 35-49.

353 **Efron, B.** 1979 Bootstrap methods: another look at the Jackknife. *Annals of Statistics*, **7**, 1-26.

- 354 **Gabriel, K.R.** 1971. The biplot graphic display of matrices with application to principal component
355 analysis. *Biometrika*, **58**, 453–467.
- 356 **Hernández-Sánchez, J., Haley, C.S. & Visscher, P.M.** 2003. Power of QTL detection using association
357 tests with family controls. *European Journal of Human Genetics*, **11**, 819–827.
- 358 **Lanzante, J.R.** 2005. A cautionary note on the use of error bars. *Journal of Climate*, **18**, 3699-3703.
- 359 **Lipcius, R.N. & Herrnkind, W.F.** 1992. Molt cycle alterations in behavior, feeding and diel rhythms of
360 a decapod crustacean, the spiny lobster *Panulirus argus*. *Marine Biology*, **68**, 241-252.
- 361 **MacNulty, D.R., Mech, L.D. & Smith, D.W.** 2007. A proposed ethogram of large-carnivore predatory
362 behavior, exemplified by the wolf. *Journal of Mammalogy*, **88**, 595-605.
- 363 **Martin, P.R. & Bateson, P.** 2007. *Measuring Behaviour: an Introductory Guide*. 3rd Edition.
364 Cambridge: Cambridge University Press.
- 365 **Martinez-Munoz, G. & Suarez, A.** 2010. Out-of-bag estimation of the optimal sample size in bagging.
366 *Pattern Recognition*, **43**, 143-152.
- 367 **Mielke, P.W. & Berry, K.J.** 2007. *Permutation Methods: a Distance Function Approach*. 2nd Edition.
368 New York: Springer.
- 369 **R Development Core Team.** 2011. *R: A Language and Environment for Statistical Computing*. Vienna:
370 R Foundation for Statistical Computing.
- 371 **Ramson, J.I. & Cade, B.S.** 2009. *Quantifying Equid Behavior: a Research Ethogram for Free-Roaming*
372 *Feral Horses*. Reston, Virginia: US Geological Survey.
- 373 **Renner, M.J. & Lussier, J.P.** 2002. Environmental enrichment for the captive spectacled bear
374 (*Tremarctos ornatus*). *Pharmacology Biochemistry and Behaviour*, **73**, 279-283.

- 375 **Tabachnick, B.G. & Fidell, L.S.** 1989. *Using Multivariate Statistics*. New York: Harper Collins.
- 376 **Underwood, A.J.** 1996. *Experiments in Ecology: Their Logical Design and Interpretation using*
377 *Analysis of Variance*. Cambridge: Cambridge University Press.
- 378 van den Berg, L., Schlider, M.B.H. & Knoll, B.W. 2003. Behaviour genetics of canine aggression:
379 Behavioural phenotyping of golden retrievers by means of an aggression test. *Behaviour Genetics*,
380 **33**, 469-483.
- 381 **Wheller, S.H. & Bennett, C.L.** 2001. Twenty-four hour activity budgets and patterns of behavior in
382 captive ocelots (*Leopardus pardalis*). *Applied Animal Behaviour Science*, **71**, 67-79.
- 383
- 384

385 Table 1. Ethogram of behaviours used for activity budget data collection of tigers

Behaviour	Description of behaviour (where required)
Eating	-
Drinking	-
Playing	Engaging in playing activities alone
Social interaction	Interacting with another tiger – either aggressive or affiliative; including grooming one another
Rolling	-
Scent marking	Spraying an object, rubbing back paws on ground or rubbing head against objects.
Walking	-
Running	-
Pacing	Repeated walking in the same pattern without an apparent goal.
Alert standing	-
Alert sitting	-
Alert laying down	Lying down with eyes open
Not alert laying down	Lying down with eyes closed
Stalking	Walking slowly with eyes fixed on one object
Grooming	-
Defecating/urinating	-
Jumping at fence	-
Vocalise	-
Other	Any behaviour that does not fit into any of the above descriptions.

386

387

388 Table 2: Ethogram of behaviours used for activity budget data collection of elephants

Behaviour	Description of behaviour (where required)
Feeding	-
Foraging	Actively searching for or extracting food items such as bark stripping
Moving	Excluding foraging
Resting	Including sleep
Socialising	Including both aggressive and affiliative behaviours
Vigilant	Elephant is standing alert
Drinking	-

389

390

391

392 Table 3: Ethogram of behaviours used for activity budget data collection of dogs (adapted from van
 393 den Berg et al., 2003).

Behaviour	Description of behaviour (if required)
Barking	-
Pulling (on lead)	-
Tail wagging	-
Growling	-
Jumping Up	-
Sitting still	-
Spinning	Dog spins in circles or changes direction frequently whilst on or off the lead.
Standing upright	-
Tail erect	-
Territorial Marking	Including urination
Approach other dogs	-
Panting	-
Whining/Whimpering	-
Yawn	-

394

395

396

397 Table 4. Ethogram of behaviours used for activity budget data collection of shore crabs

Behaviour	Description of behaviour (if required)
Claws outstretched	-
Hide	-
Pile	Piling on top of, or forcing themselves underneath other crabs
Still	-
Quick movement	$\geq 5 \text{ cm.s}^{-2}$
Slow movement	$< 5 \text{ cm.s}^{-2}$

398

399

400 Figure 1. Three dimensional principal component bubble plot with confidence radii for tiger
401 behavioural data. Bubbles represent individual tigers. Tigers 1 and 3 are males and 2 and 4 are
402 females.

403 Figure 2. Three dimensional principal component bubble plot with confidence radii for elephant
404 behavioural data. (a) Each bubble represents one of the five vasectomised bull elephants, overlap of
405 bubbles indicates no significant differences at the 95% confidence level between adjacent
406 individuals. (b) Pairwise bubble plot between the most behaviourally different elephants (as
407 determined in figure 2a) at 95% confidence – overlap between bubbles indicates no significant
408 difference. (c) Pairwise bubble plot between the most different elephants at 90% confidence, here
409 no overlap of bubbles occurs, so differences can be considered significant with 90% confidence.

410 Figure 3. Three dimensional principal component bubble plot with 95% confidence radii for dog
411 behavioural data. (a) the clustered group of three dogs on the right indicate red dogs, the three
412 remaining, non-overlapping bubbles indicate the green dogs. (b) combining the data into two cases,
413 green dogs (upper bubble) and red dogs (lower bubble) shows no overall significant difference in
414 behaviour in these classifications of dogs. (c) differences do occur at the 90% confidence level
415 between green and red dogs.

416 Figure 4. Three dimensional principal component bubble plot with 95% confidence radii for crab
417 behavioural data. Key: 1) Adult crabs in the absence of a visual predator stimulus, 2) Adult crabs in
418 the presence of visual predator stimulus, 3) Juvenile crabs in absence of visual predator stimulus, 4)
419 Juvenile crabs in the presence of visual predator stimulus.