In search of the best methods for multivariate selection analysis

Michael B. Morrissey

August 28, 2014

School of Biology, University of St Andrews

contact
    email: michael.morrissey@st-andrews.ac.uk
    phone: +44 (0) 1334 463738
    fax: +44 (0) 1334 463366
    post: Dyers Brae House
           School of Biology, University of St Andrews
           St Andrews, Fife, UK, KY16 9TH

Keywords: natural selection, selection gradients, colinearity, tables of statistics, regularised regression, projection-pursuit regression, informative priors

8 Abstract

9 1. Regression is an important method for characterising the form of natural selection from
10 individual-based data. Many kinds of regression analysis exist, but few are regularly em-
11 ployed in studies of natural selection. I provide an overview of some of the main underused
12 types of regression analysis by applying them all to test analyses of viability selection for
13 lamb traits in Soay sheep (*Ovis aries*). This exercise highlights known problems with exist-
14 ing methods, uncovers some new ones, and also reveals ways to harness underused methods
15 to get around these problems.

16 2. I first estimate selection gradients using generalised linear models, combined with recently-
17 published methods for obtaining quantitatively interpretable selection gradient estimates
18 from arbitrary regression models of trait-fitness relationships. I then also apply generalised
19 ridge regression, the lasso, and projection-pursuit regression, in each case also deriving
20 selection gradients. I compare inferences of non-linear selection by diagonalisation of the $\boldsymbol{\gamma}$
21 matrix and by projection-pursuit regression.

22 3. Selection gradient estimates generally correspond across different regression methods. Al-
23 though there is little evidence for non-linear selection in the test datasets, very problematic
24 aspects of the behaviour of analysis based on diagonalisation of the $\boldsymbol{\gamma}$ are apparent. In addi-
25 tion to better-known problems, (i) the direction and magnitude of estimated major axes of
26 quadratic selection are biased toward directions of phenotype that have little variance, and
27 (ii) the magnitudes of selection of major axes of variance-standardised $\boldsymbol{\gamma}$ are not themselves
28 interpretable in any standardised way.

29 4. While all regression-based methods for analysis of selection have useful properties, projection-
30 pursuit regression seems to stand out. This method can: (i) provide both dimensionality-
31 reduction, (ii) be the basis for inference of quantitatively interpretable selection gradients,
32 and (iii) by characterising major axes of selection, rather than of linear or quadratic selec-
33 tion separately, provide biologically-interpretable inference of non-linear selection.

34 1   Introduction

35 Understanding multivariate microevolutionary parameters is currently one of the key challenges
36 of evolutionary quantitative genetics (Blows, 2007; Philips & Arnold, 1989; Walsh & Blows,
37 2009). It is now well established that univariate and bivariate views of the genetics and selec-
38 tion of ecologically-important traits can, and perhaps even generally will, fail to reveal critical
39 aspects of microevolutionary processes, including selected axes of phenotype and genetic con-
40 straints (Dickerson, 1955; Roff & Fairbairn, 2007). Furthermore, microevolutionary parameters
41 of natural populations are likely to vary with many aspects of population structure, including
42 age and sex (Lande & Arnold, 1983; Poissant *et al.*, 2008), space (Siepielski *et al.*, 2013), time
43 (Bell, 2010; Morrissey & Hadfield, 2012; Siepielski *et al.*, 2009), and environmental conditions
44 generally (Carlson & Quinn, 2007; Grant & Grant, 2002; MacColl, 2011). Consequently, char-
45 acterisation of key aspects of the evolutionary process is generally very challenging, not only in
46 the ecological insight required to conceive data collection strategies and conduct analyses, but
47 also in that collection of required quantities of relevant data in realistic conditions and under
48 any particular regime of population structure is often very difficult. Here, I consider methods for
49 multivariate selection analysis with special focus on the biological interpretability of inferences
50 about multivariate selection from limited data. I consider viability selection of skeletal size,
51 mass, horn length, and burden of an ectoparasite in male and female Soay sheep lambs under
52 two different population dynamic regimes.

53   The best known pitfall of interpreting tables of statistical results is the problem of multiple
54 testing and false positives (Rice, 1989). Less appreciated complexities pertain to statistical
55 estimates themselves. Biological interpretation of statistical inferences about natural selection
56 generally involves consideration of tables of selection coefficients. Tables of *estimated* selection
57 coefficients will generally have very undesirable properties. Many of the aspects of selection in
58 which we may be primarily interested are not represented by individual selection coefficients,
59 but rather are obtained by applying mathematical procedures to tables of estimated selection
60 coefficients (gradients, typically). Even when applied to a table of selection coefficients that
61 are obtained by an unbiased method, few properties of tables of selection coefficients will have

62 desirable statistical properties, and in general, "doing statistics on statistics" can easily generate

63 complex statistical artefacts that can appear to represent meaningful and interesting biological

64 results.

65   A simple illustration of potential biases in interpreting tables of evolutionary quantitative

66 genetic parameters arises from the geometric interpretation of the multivariate selection gradient.

67 The length of the gradient, or its vector norm, denoted $||\boldsymbol{\beta}||$, represents an important aspect

68 of the total strength of multivariate directional selection. In multivariate studies, geometric

69 properties such as $||\boldsymbol{\beta}||$ are often integral to the best theory that we can apply to understanding

70 how selection and genetics interact to generate evolutionary trajectories (Hansen & Houle, 2008;

71 Walsh & Blows, 2009). However, the length of an estimated selection gradient vector – even

72 one composed of individually unbiased component selection gradient estimates such as those

73 generated by multiple regression analysis (Lande & Arnold, 1983; Morrissey & Sakrejda, 2013)

74 – is biased in a potentially biologically misleading way.

75   A simplified model is instructive. Consider a vector of k selection gradients, with equal

76 absolute values of b, i.e., the true value of selection for each trait (for whatever scaling of

77 phenotype has been deemed appropriate for the study) is equal. Consider that this true selection

78 gradient is estimated with error. Assume that each estimated selection gradient $\hat{\beta}_i$ is drawn

79 from a normal distribution according to $\hat{\beta}_i \sim N(b_i, s^2)$, so, each estimated gradient is unbiased,

80 sampling errors are independent, and standard errors, s, are equal across estimates. The true

81 norm of $\boldsymbol{\beta}$ is $\sqrt{kb^2}$. The expected value of the sum of squared estimated elements of $\boldsymbol{\beta}$ is $k(b^2+s^2)$,

82 as opposed to the sum of the true squared elements, which is kb^2. An exact expression for the

83 expected value of the norm of $\hat{\boldsymbol{\beta}}$ is not easily obtained. However, since $\sqrt{x}$ is a monotonic

84 function of x, it follows from $k(b^2 + s^2) \geq kb^2$ that $E[||\hat{\boldsymbol{\beta}}||]$ must be greater than $||\boldsymbol{\beta}||$, i.e.,

85 upwardly biased, whenever $s > 0$.

86   A first order approximation for $E[||\hat{\boldsymbol{\beta}}||]$ is thus $\sqrt{k(b^2 + s^2)}$, and this allows us to start to

87 get a handle on the nature of this upward bias. Bias is normally expressed as the difference

88 between the expected value of an estimator, e.g., $E[\hat{x}]$, and the true value of the estimator's

89 target estimand, x. Here, where both the true and estimated values of the length of a vector

90 must be non-negative, a proportional approach may be more intuitive. We can express $\frac{E[||\hat{\boldsymbol{\beta}}||]}{||\boldsymbol{\beta}||}$

91 (which will have a value of 1 in the absence of bias) in terms of the proportional sampling error,

92 $p = \frac{s}{b}$. Substituting pb for s in the expressions above and simplifying gives

$$\frac{E[||\hat{\boldsymbol{\beta}}||]}{||\boldsymbol{\beta}||} \approx \sqrt{1 + p^2}.$$

93 This again indicates that estimates of the length of $\boldsymbol{\beta}$, given individually unbiased component

94 elements of $\hat{\boldsymbol{\beta}}$, will be upwardly biased[1]. Furthermore, this expression illustrates that the problem

95 is severe. Since standard errors of selection gradients are generally as large as most selection

96 gradients (so $p \approx 1$; remembering that the distribution of selection gradients in the literature also

97 provides an upwardly biased impression of the average magnitude of selection; Hereford *et al.*

98 2004), upward bias in the estimated strength of multivariate directional selection on the order of

99 40% should be expected. Also, the assumptions of the instructive example should not hinder the

100 generality of its interpretation. The basic principle will hold for arbitrary distributions of true

101 values of selection gradients. Furthermore, sampling covariances among elements of $\hat{\boldsymbol{\beta}}$, as arise

102 from phenotypic covariances, will cause larger biases than the simple calculation suggests. This

103 is the principle of variance inflation under multicolinearity, which has recently been reviewed by

104 Dormann *et al.* (2013) in the context of ecological statistics.

105   The goal of this study is to explore a variety of approaches to selection analysis, in order to

106 determine what methods hold the most promise for making robust inferences of different aspects

107 of multivariate selection. I apply a range of regression methods to analyses of multivariate

108 selection of Soay sheep lamb traits, including generalised linear models, regularised generalised

109 regression models, and projection-pursuit regression. I use a recently-described approach for

110 obtaining selection gradient estimates from general fitness functions (Morrissey & Sakrejda,

111 2013) to obtain quantitative inferences of selection gradients from each of these analyses. I also

112 explore the properties of estimated major axes of quadratic selection, and of selection analysis

113 of principle components of the multivariate phenotype. These methods all provide tables of

114 selection gradients that may differ in bias, and other aspects of informativeness, with respect to

---

[1]This approximation for the proportional bias is itself somewhat upwardly biased. If selection gradients can be scaled such that their standard errors are equal to one (as a hypothetical instructive situation), the expected norm of the estimated selection gradient vector is given by the expectation of a chi distribution. This does not lead to a simple informative expression, but numerical analysis shows that the approximation $\frac{E[||\hat{\boldsymbol{\beta}}||]}{||\boldsymbol{\beta}||} \approx \sqrt{1 + p^2}$ upwardly estimates the bias by about 10% for k and $p \approx 0.7$, and is otherwise a close depiction of bias.

115   different aspects of multivariate selection.

2   Methods and Results

2.1   Example study system and data

118   Soay sheep on Hirta, St Kilda, in the Outer Hebrides, have been monitored in an individual-based
119   study since 1984 (Clutton-Brock & Pemberton, 2004). The portion of the phenotypic data used
120   here are collected each August, when a large portion of the Soay sheep resident in the Village
121   Bay study area are captured. I analyse body mass (kg), hind leg length (mm), horn length
122   (mm), and number of keds (*Melophagus ovinus*), an ectoparasite, of lambs measured in August.
123   Aspects of size have previously been shown to be related to survival (e.g., Clutton-Brock *et al.*
124   1992; Milner *et al.* 1999), and to have complex phenotypic and genetic covariances with lifetime
125   fitness (Morrissey *et al.*, 2012a). Horn size is also closely related to aspects of both survival and
126   reproduction (Coltman *et al.*, 1999; Johnston *et al.*, 2013; Robinson *et al.*, 2006). Although keds
127   cause some skin irritation (Wilson *et al.*, 2004) their presence or prevalence has not previously
128   been related to fitness, and this parasite does not appear to impact negatively on other aspects
129   of sheep performance. I focus on traits in lambs, and furthermore restrict the dataset to those
130   individuals with the normal horn morph (Clutton-Brock & Pemberton, 2004).

131   The month, and usually day, of death is known for nearly all individuals, allowing us to
132   determine viability from the time of measurement in August through to one year of age, defined
133   operationally as 1st April in the year following birth. The Soay sheep population experiences a
134   wide range of over-winter survival rates, with pronounced crashes in some years (Clutton-Brock &
135   Pemberton, 2004). For all selection analyses, I therefore divide the dataset into four subsets, for
136   male and female lambs in crash and non-crash years. Cohorts born in springs prior to overwinter
137   crashes are: 1988, 1991, 1994, 1997, 2004 and 2011. Sample sizes and mean survival rates are:
138   males in non-crash years: $n = 633$, $\bar{W} = 0.687$, females in non-crash years: $n = 213$, $\bar{W} = 0.803$,
139   males in crash years: $n = 281$, $\bar{W} = 0.359$, and females in crash years: $n = 117$, $\bar{W} = 0.470$.
140   Sample size is smaller for females because the expression of the horn polymorphism is sex-specific,
141   and fewer females have normal horns (Clutton-Brock & Pemberton, 2004). Means, variances,

142 and correlations among traits in each sex and environmental condition are given in table 1. All

143 traits, i.e., mass, leg length, horn length, and log ked number, were standardised to unit variance

144 within each of the four datasets.

2.2 General strategy for selection gradient estimation

146 Analyses in sections 2.3 - 2.7 all use a common framework for selection gradient estimation.

147 In each case, the relationship between multivariate phenotype and expected individual fitness,

148 $E[W_i] = f(\mathbf{z}_i)$, is first determined using a generalised regression model. Subsequently, pop-

149 ulation mean fitness, given the sample of phenotypes $\mathbf{z}$ and the function $f(\mathbf{z})$ is obtained by

150 $\bar{W} = \frac{1}{n} \sum_i^n f(\mathbf{z}_i)$. The first and second partial derivatives of population mean fitness with re-

151 spect to population mean phenotype are then calculated by numerical methods. When scaled

152 by dividing by population mean fitness, these derivatives provide estimates of directional and

153 quadratic/correlational selection gradients, i.e., $\beta_i = \frac{\delta \bar{W}}{\delta \bar{z}_i} \bar{W}^{-1}$, and $\gamma_{i,j} = \frac{\delta^2 \bar{W}}{\delta \bar{z}_i \delta \bar{z}_j} \bar{W}^{-1}$, respectively.

154 This method is described in detail in Morrissey & Sakrejda (2013). Where appropriate, stan-

155 dard errors were calculated and statistical hypothesis tests were applied using the parametric

156 bootstrap method also described in Morrissey & Sakrejda (2013). All traits were standardised

157 to unit variance.

158 The Morrissey & Sakrejda (2013) method for obtaining selection gradients estimates from

159 arbitrary inferences of $E[W_i] = f(\mathbf{z}_i)$, directly calculates the partial derivatives of population

160 mean fitness with respect to population mean phenotype, scaled to the relative fitness. These

161 quantities are returned regardless of the distribution of phenotype. For example, if the distri-

162 bution of one or more traits is skewed, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ will still reflect the first and

163 second partial derivatives of mean relative fitness with respect to phenotype. This definition of

164 $\boldsymbol{\beta}$ relates to evolutionary change via $\Delta \bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}$ under the assumption that breeding values are

165 multivariate normal (see proof in appendix). In contrast, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ provided

166 by the familiar regression analysis described by equation 16 in Lande & Arnold (1983), will not

167 predict evolutionary change when the phenotype is not multivariate normal, even if breeding

168 values are MVN.

169 The shape of $f(\mathbf{z}_i)$, as obtained by generalised regression analysis, will be determined in part

170   by the link function. If $f(\mathbf{z}_i)$ is a linear function on the linear predictor scale, i.e., takes the form of

171    $E[W_i] = link^{-1}(\mu + \sum_j^k b_j z_{i,j})$, then the curvature will be entirely determined by the shape of the

172   link function. Estimates of $\boldsymbol{\beta}$ obtained from such a model of the fitness function will generally

173   provide robust inference of directional selection, but estimates of $\boldsymbol{\gamma}$ should not generally be

174   interpreted biologically. When quadratic, or otherwise curved (e.g., spline) generalised regression

175   models are used for $f(\mathbf{z}_i)$, the link function will generally have very little effect on estimates of

176   either $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$. For example, models of binary outcomes (e.g., survival) could equally be fitted

177   using logit or probit link functions. For any given dataset, the parameters of $f(\mathbf{z}_i)$ will differ

178   between models using the logit and probit link functions, but the shape of $f(\mathbf{z}_i)$ on the expected

179   fitness scale, and therefore estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, will typically differ trivially.

180   All analyses in the present work consider relatively simple distributions of fitness. In particu-

181   lar, all analyses and empirical examples involve a binary (survival) fitness response. The method-

182   ological focus is thus on aspects of inferring selection of the multivariate phenotype. These issues

183   should be seen as complimentary to other ongoing avenues for methodological development of

184   methods for the analysis of natural selection. The work here is hopefully complimentary to,

185   for example, methods for using information about the life cycle to construct sensible models of

186   variation in fitness (Geyer *et al.*, 2007; Shaw & Geyer, 2010), efforts to characterise selection in

187   a demographic context (Engen & Saether, 2014; Engen *et al.*, 2012; Morrissey *et al.*, 2012b), and

188   application of theory to disentangle purely correlative from direct and indirect effects of traits

189   on fitness (Morrissey, 2014).

190   **2.3   Selection differentials and multiple regression-based estimation of selection gra-**

191   **dients**

192   I obtained variance-standardised directional selection differentials S for each trait by calculating

193   the difference between mean phenotype weighted by fitness and mean phenotype before selection.

194   I obtained quadratic selection differentials as $\mathbf{C} = \Delta\mathbf{P} + \mathbf{SS}^T$, where $\Delta\mathbf{P}$ is a matrix of trait

195   variances and covariances, weighted by relative fitness, minus the variance and covariances before

196   selection, and $\mathbf{S}$ is the vector of directional selection differentials. I obtained standard errors and

197   p-values of selection differentials by case bootstrapping.

198 I obtained standardised directional and quadratic selection gradients by first fitting gener-

199 alised linear models (glm) with a binomial responses, using the R package MGCV (Wood, 2006),

200 and (linear predictor scale) linear, quadratic, and interaction effects for all traits and trait combi-

201 nations, for each of the four datasets. I then obtained the selection gradient estimates from these

202 fitted models (see section 2.2), as implemented in the R package GSG (Morrissey & Sakrejda,

203 2013), with standard errors and p-values, using a parametric bootstrap.

204 Survival covaries positively with mass and leg length in both sexes and in both environmental

205 conditions (table 2a). In crash years only, but in both males and females, survival covaries

206 positively with horn size as well. No consistent patterns occur in changes in variances and

207 covariances due to selection, over and above those necessarily associated with changes in the mean

208 (Endler, 1986; Lande & Arnold, 1983), with the exceptions of some marginally non-significant

209 values, and one nominally significant value (i.e., without accounting for multiple tests) for the

210 change in the covariance of horn length and ked number.

211 Selection gradients revealed that covariance of survival with mass and leg length is primarily

212 directly attributable to variation in mass in non-crash years (table 2b). Also in non-crash

213 years, horn length has negative direct effects on survival, again in both sexes, i.e., the slightly

214 positive and non-significant covariances of horn length and survival arise via opposite effects of

215 correlated selection of mass, and direct selection of horn length. Inference of the direct causal

216 structure of selection in crash years appeared to be hindered in part by smaller sample sizes

217 for crash years, compared with the relatively high degree of correlation of phenotypic traits

218 (which happened across conditions; table 1). Importantly, this should not be taken as a lack of

219 statistically significant selection: the covariances of traits and fitness (table 2a), arise somehow,

220 and multiple regression analysis can only attribute this covariance to the traits that are included

221 in the analysis. Thus there is significant selection, but there is also a statistical failure to

222 robustly partition total selection into direct effects among the available predictor variables. This

223 alone is an important property of a table of multiple regression coefficients, and potentially an

224 interpretive trap (see also discussion in Mitchell-Olds & Shaw 1987); non-significance of each

225 gradient in isolation (table 2b) does not correspond to non-significance of total selection (table

226 2a).

227 **2.4   Major axes of the quadratic approximation of the fitness surface**

228 I further investigated multivariate quadratic selection following methods discussed and promoted

229 by Philips & Arnold (1989) and by Blows (2007). To characterise the major orthogonal axes of

230 quadratic selection, I performed canonical rotations of each matrix of estimated quadratic and

231 correlational selection gradients

$$\boldsymbol{\gamma} = \mathbf{M\Lambda M}^T \tag{1}$$

232 where $\mathbf{M}$ is a matrix of orthogonal eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the

233 associated eigenvalues. Values in $\boldsymbol{\Lambda}$ are interpreted as the quadratic selection gradients of the

234 new independent axes of the quadratic component of the relative fitness surface.

235   I constructed null distributions of the magnitudes of the eigenvalues of the rotated $\boldsymbol{\gamma}$ matrices

236 using an algorithm very similar to that suggested by Reynolds *et al.* (2010). I first generated 1000

237 datasets with the original phenotypic data (separately for each combination of sex and crash vs.

238 non-crash conditions) and permuted values of fitness. I then re-fitted the multivariate quadratic

239 logistic regression model, and for each logistic model fitted to the permuted fitness data, I re-

240 calculated the associated selection gradients, as above. From each set of selection gradients for

241 each permuted dataset, I rotated the $\boldsymbol{\gamma}$ matrix and recorded the eigenvalues (i.e., the quadratic

242 selection gradients of the diagonalised estimated $\boldsymbol{\gamma}$ matrix), ordered by their absolute values.

243 Statistical hypotheses tests associated with the comparison of observed values to permuted values

244 are given in table 3.

245   Some authors have reported statistical hypothesis tests of selection along axes with smaller

246 eigenvalues. While there is potentially some value in considering statistical hypothesis tests of

247 minor axes, when larger axes are non-significant, it is not clear that any interpretive gain could

248 outweigh the dangers of multiple testing. In the present analyses, across 16 tests of four axes of

249 quadratic selection, in each of both sexes and both crash and non-crash years, no permutation-

250 based tests of any axis were statistically significant at a marginal value of 0.05 (table 1).

251   The major axis of the diagonalised $\boldsymbol{\gamma}$ matrix in both sexes and in both environments involved

252 loadings of mass and leg length in opposite directions (table 1). In other words, the main axis of

253 estimated selection was aligned in the direction of phenotype that had the least variance. This

254 is probably an artefact of the fact that selection is most difficult to characterise in this direction,

255 and therefore sampling error will produce the largest errors in the direction of phenotype with

256 the least variance. A second interpretive difficulty is apparent in figure 1. Even though the

257 analysis is conducted on unit variance-standardised values of phenotype, the major axes of $\boldsymbol{\gamma}$

258 cannot be interpreted with the benefits that come from variance standardisation. Despite the

259 fact that the first axes represent much greater absolute curvature than the second axes in each

260 case (table 3), the amount of variation in fitness associated with the first two axes - over the

261 distribution of phenotype in those directions - is very similar in two cases (figure 1a,c), and in

262 two cases the variance in fitness associated with the second estimated axis is clearly greater that

263 that associated with the first (figure 1b,d).

264 ## 2.5   Regularised regression-based selection gradient estimates

265 Elastic net regularisation (Zou & Hastie, 2005) is a general form of biased regression estimation

266 that includes ridge regression (Tikhonov & Arsenin, 1977) and least absolute shrinkage and

267 selection operator ("the lasso"; Tibshirani 1996) as special cases. Where least squares regression

268 obtains estimated regression coefficients $\boldsymbol{b}$ by minimising $||\boldsymbol{y} - \mathbf{X}\boldsymbol{b}||^2$, the elastic net minimizes

269 $(||\boldsymbol{y} - \mathbf{X}\boldsymbol{b}||^2 + \alpha||\boldsymbol{b}||^2 + (1 - \alpha)||\boldsymbol{b}||)$. When $\alpha = 1$, the analysis is a ridge regression, and when

270 $\alpha = 0$, the analysis is the lasso.

271   Both ridge regression and the lasso thus minimise penalised sums of squares, with the goal

272 of maximising predictive ability, rather than fit to the sample data. In practice, ridge regression

273 reduces the overall magnitude of regression coefficients, relative to least-squares regression, and in

274 particular, gives more plausible values for regression coefficients associated with highly correlated

275 predictor variables. The lasso also produces shrunken values, but will generally shrink different

276 coefficients to a much greater extent, in particular, potentially assigning zero values to coefficients

277 associated with variables that have no probable predictive ability. Ridge regression and the lasso

278 therefore have properties that may be desirable overall, and that can be particularly desirable

279 when predictors are highly correlated, as is often the case in selection analysis.

280   I used generalised elastic net (with ridge regression and the lasso, and a combination of

281 the two with $\alpha = 0.5$) regression to estimate selection gradients, as above, by first estimating

282 fitness functions, and then obtaining selection gradient estimates from those functions. I used

283 the function `cv.glmnet()` in the R package GLMNET (Friedman *et al.*, 2008) to fit the ridge

284 regression, lasso and elastic net regressions ($\alpha = 0.5$) with binomial responses by generalised

285 cross-validation, and used those estimated regression coefficients based on the penalty parameter

286 λ that minimised the cross-validation score. All estimated selection gradients derived from these

287 models of the fitness function are given in table 4.

288   In non-crash years, results of lasso, ridge, and elastic net regressions yielded selection gradients

289 (table 4a) that largely match gradients obtained from glm-based inferences of the fitness functions

290 (table 2b). Gradient estimates that are near zero and not statistically significant in glm-based

291 analysis are often shrunken to zero or very near zero by the lasso, and substantially shrunken

292 by the ridge regression, with the elastic net yielding intermediate results.

293   When applied to data from crash years, where partitioning of direct effects proved more

294 difficult in the glm-based analysis, the regularised regression yielded inferences that may be

295 somewhat more useful. For example, mass was identified as being under positive selection.

296 It does not make sense to try to obtain standard errors or p-values, for example, using the

297 bootstrap, as above, for regularised regression analyses. To some extent, the "significance" of

298 each coefficient is represented in its estimated value, in the degree to which it is shrunken,

299 especially for coefficients with non-zero values in lasso regression. For sequential model-building

300 exercises, new experimental methods can provide p-values for the lasso (Lockhart *et al.*, 2013).

301   As a visual measure of the total strength of directional and quadratic multivariate selection,

302 I predicted expected absolute fitness (survival) for each individual from the fitted glm and ridge

303 regression models. The distributions of expected absolute fitness are shown in figure 2. This

304 provides a overall picture of the amount of variation in fitness that is associated with regression-

305 based inference about selection. The distributions of expected fitness from the glm, suggests that

306 on the basis of just four traits out of the entire multivariate phenotype, one could essentially

307 predict death or survival for many individuals with near certainty. On the other hand, the ridge

308 regression represents a seemingly more appropriately modest inference of the predictive power

309 of a handful of traits. This does not demonstrate that the non-regularised regression analysis is

310 somehow wrong; rather, it is another way of illustrating ways in which alternative methods may

311 have more reasonable interpretations for some purposes.

2.6   Selection of major axes of P

313 An alternative and common (e.g., Bolnick & Lau 2008; Grether 1996; Schluter & Smith 1986)

314 means of reducing the dimensionality of a selection analysis is to consider only the relationship

315 between major axes of phenotypic variation and fitness. For each dataset, I applied a spectral

316 decomposition of the phenotypic correlation matrix, and then rotated the phenotypic data onto

317 the two largest axes (largest eigenvalues) of $\mathbf{P}$. Specifically, given the first two eigenvectors

318 of the distribution of phenotype, $\mathbf{L}_2$, and the original phenotypic records $\mathbf{z}$, the new 'traits'

319 representing loadings of the first two major axes of phenotype are $\mathbf{z}_2 = \mathbf{z}\mathbf{L}_2$. I then estimated

320 selection gradients of $\mathbf{z}_2$ by first fitting a glm with linear, quadratic and interaction terms, and

321 then obtained selection gradients from this function, as above, and also generated standard errors

322 and applied statistical hypothesis tests using the parametric bootstrap. Variance-standardised

323 selection gradient estimates pertaining to the major axes of $\mathbf{P}$ are given in table 5. These

324 estimates are interpretable as the selection intensities of the main axes of phenotype as defined

325 by the correlation structure of the traits. A variety of other standardisations are possible. Each

326 would require different interpretation, and each may reveal different information about natural

327 selection.

328   In these datasets, estimating selection of compound axes of phenotype does not provide very

329 meaningful inference of multivariate selection. In the example analyses, this practice revealed a

330 pattern of "bigger is better" across all traits, i.e., there would appear to be positive selection of

331 an axis onto which all three of the morphometric traits load positively. This fails to elucidate

332 patterns that are otherwise easily obtained (tables 2 and 4). In particular, the "bigger is better"

333 result that arises from analysis of principle components of phenotype conflicts with two important

334 findings: (i) Mass, rather than structural size is more proximally related to fitness, certainly in

335 non-crash years, and probably overall, and (ii) while large horns appear to be positively selected

336 via their loading of the directionally selected first 'size' axis of phenotype, horns are probably

337 either detrimental or unrelated to lamb survival in most circumstances (tables 2 and 4).

338 **2.7 (Generalised) projection-pursuit regression-based selection gradients and fit-**
339 **ness surface estimation**

340 The use of projection-pursuit regression (Friedman & Stuetzle, 1981) to estimate fitness functions
341 has been little-used since its introduction to the field by Schluter & Nychka (1994). This method
342 reduces the dimensionality of the problem by seeking the orthogonal axes of the multivariate
343 phenotype that maximise the explained variation in fitness. Each axis is characterised by a ridge
344 function, typically characterised by a semi-parametric smooth regression function. Briefly, the
345 response variable (or its linear predictor) is modelled as $link(E[y]) = \mu + \sum_i^k f_i(\mathbf{b}_i X) + e_i$, where
346 $f_i()$ are the ridge functions associated with the estimated axes of phentoype that best explain y,
347 as defined by $\mathbf{b}$; in the notation employed here for fitness functions, $y = W$ and $\mathbf{X} = \mathbf{z}$. Both $\mathbf{b}$
348 and the parameters of the arbitrary ridge functions $f_i()$ are estimated, simultaneously yielding
349 inference of the axes of phenotype that are selected, and of the form of selection. A complete
350 description of the method, with specific application to inference of fitness functions, is given in
351 Schluter & Nychka (1994).

352 I implemented a generalised projection-pursuit function (gppr) by wrapping the function
353 `ppr()`, in the R package STATS in an iterative re-weighting function. I used cubic regression
354 spline regressions fitted by generalised cross validation for the ridge functions (matching the
355 implementation by Schluter & Nychka 1994). I characterised each fitness function with gppr
356 functions with one and two main axes. As above, I extracted selection gradient estimates from
357 the inferred fitness functions following Morrissey & Sakrejda (2013). The gppr function, `gppr()`,
358 and a function to obtain selection gradients from gppr analysis, `gppr.gradients()`, are included
359 in version 2.0 of the R package `gsg` (originally described in Morrissey & Sakrejda 2013). I
360 obtained estimates of the selection gradients of the major axes of selection as determined by
361 gppr by rotating the phenotypic data onto the axes identified in the gppr analysis, and then
362 refitting the model using `gam()` in `mgcv`, with univariate splines for each axis. I then recovered
363 the selection gradients of these axes using `gam.gradients()` in `gsg`.

364 Because familiar hypothesis testing is not directly compatible with models fitted by cross-
365 validation, I applied a randomisation procedure to help give an idea of how much variation was

366 explained by the gppr models, over and above statistical noise. I made 1000 datasets for each
367 sex and environmental condition, each with the available sample size and observed distribution
368 of phenotype, but with randomised survival records. I then applied the gppr analyses with
369 one and two ridge functions, predicted individual absolute fitness for each fit, and recorded the
370 variance in predicted absolute fitness for each fitted function for each randomised dataset. I then
371 compared the variance in predicted absolute fitness, and the differences in predicted absolute
372 fitness between models with one and two dimensions, between the randomised datasets and the
373 real datasets.

374   The gppr analyses revealed largely directional and linear selection (figure 3). The only sug-
375 gestion of curvature of the major axes of selection was for females in non-crash years, and is more
376 interpretable as expected fitness asymptotically approaching one, rather than any mechanism of
377 non-linear selection. Because selection appears to be largely linear, the loadings of phenotype
378 onto the major axis of selection closely matched estimated directional selection gradients (table
379 2b).

380   In all cases, one axis of phenotype explained substantial variation in fitness (table 6). For all
381 analyses, I plotted the first major axis for ease of interpretation (figure 3); for males in non-crash
382 years, the predictions based on two axes were not interpretable in terms of any simple pattern
383 of selection. The gppr functions with two dimensions of selection did not explain much more
384 variation than the replicated analyses of randomised datasets (table 6), except for males in non-
385 crash conditions; however, in that case, the amount of additional explained variation associated
386 with the two dimension model was nonetheless modest.

2.8   Supplementary simulations

388 For better or worse, the primary approach in this study was to compare the inferences that
389 could be made by applying different types of regression-based selection analyses to the same
390 empirical datasets. For better, consideration of the behaviour of the different analyses in their
391 application to real data has revealed a number of phenomena that might not otherwise have
392 surfaced. For worse, it is rarely possible to determine with certainty which analyses are most
393 likely to best reflect reality when conducting a case study on real data. Also, it is not necessarily

394 clear whether and which phenomena that have occurred in the specific analyses here will be
395 important in general. While the purpose here is not to conduct any comprehensive simulation
396 studies, two specific issues seem to necessitate a little further investigation. These and other
397 issues would certainly benefit from more comprehensive studies.

398 ### 2.8.1 Regularised regression analyses

399 The application of ridge regression, the lasso, and the elastic net regression to the Soay sheep
400 lamb datasets did not reveal any major benefits relative to other methods. One reason for this
401 may be that the dimensionality of the selection analysis problems in this study are rather low,
402 i.e., four traits. Combined with the fact that the ecological relevance of each trait is reasonably
403 intuitive to a human, we are inclined to think about selection on a trait-by-trait basis. It would be
404 a shame if the potential benefits of these analyses were marginalised because their benefits are not
405 immediately apparent in a single case study. One major potential benefit of regularised regression
406 is that it should be expected to provide some reduction in the tendency for statistical noise to
407 generate biases in some geometric properties of selection gradients, in particular, in the total
408 length of $\hat{\boldsymbol{\beta}}$. If evolutionary quantitative genetic studies are able to become more multivariate,
409 and are able to apply geometric concepts to understanding evolution, as advocated for example
410 by Blows (2007) and Walsh & Blows (2009), robust inference of geometric properties of quantities
411 such as selection gradient vectors and $\mathbf{G}$ matrices will become increasingly important. Here, I
412 continue with the idea from the introduction that the norm of $\hat{\boldsymbol{\beta}}$ may be a very biased estimator
413 of $||\boldsymbol{\beta}||$, and test more generally by simulation whether regularised regression can provide better
414 inference. In geometric interpretations of microevolutionary parameters, $||\boldsymbol{\beta}||$ is just one of
415 several important geometric quantities, appearing, for example, in theoretically well-justified
416 metrics of evolutionary constraint (Hansen & Houle, 2008).
417   I simulated 24 different scenarios, including every combination of: (i) sample sizes of n=100,
418 200, and 500 individuals, (ii) number of trait, $k = 4, 10$, (iii) normal (μ=0, σ=0.5) and t-
419 distributed (μ=0, σ=0.5, df=1) logistic scale regression gradients of expected fitness (i.e., W
420 in [0,1]), and (iv) low and high dispersion of eigenvalues of the $\mathbf{P}$ matrix. For each simulation,
421 unique $\mathbf{P}$ matrices were simulated from an inverse Wishart distribution of V=$\mathbf{I}$, and $\nu = 20$

422 and 5, when $k = 4$ for low and high dispersion of $\mathbf{P}$, respectively, and $\nu = 30$ and 11, when

423 $k = 10$ for low and high dispersion of $\mathbf{P}$. Each $\mathbf{P}$ matrix was standardised to unit variance

424 in each trait. For each simulation, I simulated n records of k traits, $\mathbf{z}$, with mean vectors of

425 $\mathbf{0}$, and covariance $\mathbf{P}$. For each simulation I also drew unique logistic scale gradients of fitness

426 with respect to phenotype, $\mathbf{b}$, according to either the normal or t-distribution, depending on the

427 scenario. I then simulated individual fitness records from a binomial distribution with expected

428 value $logit^{-1}(\mathbf{zb}')$.

429   I calculated the resulting true selection gradients by a modification of the Morrissey & Sakre-

430 jda (2013) algorithm. I generated 10^6 records of phenotype according to the true value of $\mathbf{P}$

431 for each replicate simulation. For each phenotypic record, I then calculated expected absolute

432 fitness, and averaged these to obtain population mean expected fitness. Then, separately for

433 each of the k traits, I re-calculated population mean fitness for a modified dataset in which 0.03

434 had been added to each phenotypic record for a given trait, and repeated with subtracting 0.03.

435 I then calculated the partial derivative of population mean fitness with respect to phenotype,

436 scaled to relative fitness, i.e., the selection gradient, for each trait, by finite differences. I re-

437 peated this algorithm five times for each replicate of each simulation scenario. Values of the true

438 values of $\boldsymbol{\beta}$ agreed to the 4th decimal place in most replicate applications of the MC procedure.

439 I took the means across the five simulations to be the true values of $\boldsymbol{\beta}$.

440   I then applied logistic regression analyses with linear terms only in each replicate simulation

441 scenario. I fitted GLMs, as well as ridge, lasso, and elastic ridge regressions, and in each case

442 obtained selection gradient estimates, as described above for the case studies in Soay sheep. In

443 each replicate simulation I calculated $\frac{||\hat{\boldsymbol{\beta}}||}{||\boldsymbol{\beta}||}$ for each of the four regression-based estimates of $\hat{\boldsymbol{\beta}}$. As

444 predicted by the simple theory developed in the introduction and the appendix, the estimates

445 of the total magnitude of selection are upwardly biased in analyses non-regularised analyses,

446 especially when the $\mathbf{P}$ matrix is ill-conditioned (figure 4). Regularised regression analyses yield

447 somewhat negatively, but less, biased inference of $||\boldsymbol{\beta}||$. The analyses also reveal that the best

448 route to robust inference is to collect sufficient quantities of data (figure 4).

449 **2.8.2   Major axes of selection**

450 I further investigated the degree to which the inference of the major axes of estimated $\boldsymbol{\gamma}$ matrices

451 is biased toward minor axes of the phenotype by simulating eight different scenarios of bivariate

452 quadratic selection analyses, involving high and modest phenotypic correlations ($r = 0.8$ and

453 0.5), two sample sizes ($n = 250$ and 500), and for no selection, and for true stabilising selection

454 on one trait (fitness function was binomial with $logit(E(W|z_i)) = -0.3z_i^2$). The simulations

455 with true stabilising selection thus had a true angle between the major axes of $\mathbf{P}$ and $\boldsymbol{\gamma}$ of 45

456 degrees.

457   The null distribution of the direction of the major axis of quadratic selection, relative to the

458 major axis of phenotype, is highly biased toward orthogonality, especially when there are strong

459 phenotypic correlations between traits (figure 5a,c). Even when selection occurs, substantial

460 bias of estimated major axes of selection occurs (figure 5b). In the best case scenario, with sub-

461 stantial (true) selection which is easily characterised because of modest phenotypic correlations,

462 inferences based on rotation of estimated $\boldsymbol{\gamma}$ matrices can become robust (figure 5d). In the

463 context of multivariate selection analyses, even when correlations as high as 0.8 do not occur,

464 similarly minor axes of phenotype to the simulations in figure 5a,b are common. Consequently,

465 it seems that substantial biases in the direction of major axes of $\boldsymbol{\gamma}$ relative to $\mathbf{P}$ are likely to be

466 prevalent in most studies of multivariate non-linear selection.

467 # 3   Discussion

468 Exploration of a broad range of regression analyses for quantitative inferences of natural selec-

469 tion revealed useful properties of several under-used approaches, and also illustrated important

470 limitations of some commonly applied analyses. In particular, two previously unacknowledged

471 properties of diagonalised quadratic selection matrices seriously hinder interpretation: (i) the

472 method is biased toward apparent detection of quadratic selection that is orthogonal to major

473 axes of variation, and (ii) biases aside, the magnitudes of the major axes of variance-standardised

474 $\boldsymbol{\gamma}$ cannot be interpreted in variance-standardised units. However, in conjunction with recently

475 developed methods for quantitative inference of selection gradients from arbitrary fitness func-

476 tions (Morrissey & Sakrejda, 2013), projection-pursuit regression may be able to serve for the

477 main biological questions for which diagonalisation of $\boldsymbol{\gamma}$ has been suggested, and will also pro-

478 vide other useful properties. More generally, quantitative inference of selection gradients from

479 arbitrary regression analyses will allow useful properties of a greater range of types of selection

480 analysis to be exploited for the study of natural selection.

481   In the example datasets, several conclusions are repeatedly supported by different analyses.

482 First, while larger size is generally selected, mass is more directly associated with first winter

483 survival than is structural size, as represented here by leg length. The positive direct selection

484 of mass reported here does not contradict the previously reported associations of limb length

485 with neonatal survival Coltman *et al.* (1999), as the lack of direct effect characterised by the

486 selection differential (low here for leg length; table 2b) neither precludes association (table 2a),

487 or an indirect mechanistic effect of leg length on survival. The strong direct effect of mass is

488 supported by all regression-based inferences of selection gradients, although the pattern is more

489 tenuous in crash rather than non-crash years, and also coincides with extensive work showing

490 associations of mass with life history in this population (Milner *et al.*, 1999). While it is probably

491 not possible to conclude that mass-fitness relationships, in adults at least, are causal (Morrissey

492 *et al.*, 2012a), it seems plausible that energy reserves in lambs could well be a key causal variable

493 in determining first winter survival. It appears that for crash years, smaller available sample

494 sizes combined with high covariances among traits conspire to make separation of direct and

495 indirect effects very difficult. The regularised regression analyses, particularly the lasso, which

496 allows some degree of inherent variable selection, suggest that the pattern of mass, rather than

497 structural size, being most proximal to survival, may hold in crash years as well (table 4).

498   Whereas there is either little selection, or positive selection, *of* horn size in lambs across sexes

499 and environmental conditions (i.e., near zero or positive selection differentials), there appears to

500 be substantial selection *for* (terminology for association vs. effect following Sober 1984) smaller

501 horn size in non-crash years. The pattern in non-crash years, at least, is simple and intuitive, as

502 investment in horns is unlikely to positively influence survival in general (Johnston *et al.*, 2013;

503 Robinson *et al.*, 2006) or survival of lambs in particular. I am hesitant to interpret estimated

504 selection gradients of horn length in crash years as indicative of any environmentally-induced

505   variation in selection, as uncertainty in their estimation precludes rejection of the possibility of

506   similar selection across environmental conditions. If indeed selection does directly favour horn

507   size, at least directly with reference to the traits studied here, this could reflect utility of horns

508   for competition for scarce resources.

509   The simple "bigger is better" pattern apparently revealed by positive selection of the first

510   axis on phenotypic variation gives an impression of simple directional selection on the traits

511   underlying the first axis (mass, leg length, and horn length), in both sexes and environmental

512   conditions. All other analyses show that this pattern does not reflect reality, at least not in

513   non-crash years where direct effects of traits can be estimated with relative precision. Certainly,

514   situations will arise where principle components will reflect ecologically-relevant axes of varia-

515   tion. However, the analyses here highlight that statistically dominant and ecologically important

516   axes of variation may be very different. At the very least, results of selection analyses of prin-

517   ciple components should be approached with caution, especially when the motivation for using

518   principle components is statistical (dimensionality reduction), rather than biological.

519   Regularised regression methods (i.e., the lasso, ridge regression, and the elastic net) generally

520   supported patterns in selection gradients that were obtained from more traditional regression-

521   based analyses. The lasso may have provided improved inferences in cases where trait covariances

522   otherwise precluded inference of selection gradients. For example, it is useful that the lasso

523   was able to identify a most-probable proximate effect of mass on survival in males in crash

524   years, where other methods were essentially unable to distinguish among potential effects of the

525   different traits. Similarly, the total amounts of variation in survival that are apparently explained

526   by regression analyses (figure 2) are much more plausible for ridge, rather than for (unpenalised)

527   least-squares regression. While the application of these regularised regression analyses did not

528   greatly help interpretation of the Soay sheep example data, it is possible that they could be quite

529   useful in other circumstances, especially for making geometric interpretations about multivariate

530   selection (figure 4).

531   Two major features of the analyses of multivariate non-linear selection in the Soay sheep lamb

532   datasets highlight the difficulties in interpretation of the major axes of the quadratic selection

533   gradient matrix, $\boldsymbol{\gamma}$. It may initially seem quite bold to criticise existing methods for analysis

534 of non-linear selection based on example analyses of datasets that do not, it turns out, seem
535 to contain much non-linear selection. However, the undesirable behaviours of inferences about
536 the *estimated* γ matrix here will exist in any analysis, regardless of the underlying reality. The
537 first serious problem is that statistical noise has a very insidious effect on the orientation of the
538 *estimated* major axes of γ. The true curvature of γ will be hardest to estimate in directions
539 within $\mathbf{P}$ that have the least variance. Therefore, major axes of γ are likely to correspond to
540 minor axes of phenotype, purely as an artefact of the fact that statistical noise will create the
541 greatest estimated values in directions within $\mathbf{P}$ that have the least variance. Such a pattern
542 unfortunately has a very tempting biological interpretation, i.e., that quadratic selection and
543 multivariate phenotype are aligned. This problem is quite intuitive once one starts to consider
544 the effect of noise on a table of estimated selection coefficients such as γ. Note that this problem
545 will affect all axes of estimated γ matrices, influencing both shape and orientation. Even where
546 axes exist that are subject to quadratic selection, inference of their orientation will be hindered
547 by the fact that the orientation of other axes is biased by the shape of $\mathbf{P}$, combined with the
548 constraint of orthogonality inherent to diagonalisation (table 5). This problem arises because
549 of different amounts of variation in different directions of phenotypic space. Therefore, it will
550 affect analyses of major axes of γ matrices under any system of trait standardisation.

551   A second difficulty with spectral decomposition of γ is specific to analysis of variance-
552 standardised γ matrices. Where the original gradients are interpretable as the direct components
553 of selection intensities, i.e., they reflect the amount of fitness variation directly associated with
554 (quadratic) selection of the traits, given the standing variation in the traits, the major axes of
555 gamma do not have this interpretation. This phenomenon is particularly noticeable in figure 1b.
556 Here, the most curved direction of γ is aligned with an axis of $\mathbf{P}$ (table 1) that has little variation
557 (whether this is real, or chance, is not immediately relevant to this second point). Consequently,
558 the first axis of stabilising selection is actually associated with less variation in fitness than the
559 second axis! This is apparent in figure 1b, where the curvature of the second axis is indeed
560 less than that of the first, but it represents stronger selection because it is associated with more
561 phenotypic variation.

562   Of course, understanding multivariate selection, including multivariate non-linear selection,

563 remains extremely important. Fortunately, a variety of features of projection-pursuit regression
564 make it highly amenable to the study of multivariate selection. In combination with methods to
565 obtain quantitatively interpretable selection gradients from projection-pursuit regression-based
566 inferences of fitness functions, as applied here, this method can probably supplant the practice
567 of diagonalisation of $\boldsymbol{\gamma}$. The first major empirical benefit of projection-pursuit regression is that
568 it can be used to seek the major axes of selection, not just the major axes of directional or of
569 quadratic selection. In addition to the issues already discussed about diagonalisation of $\boldsymbol{\gamma}$, there
570 has never been any real resolution to the fact that quadratic univariate or multivariate selection,
571 considered either in isolation or in conjunction with $\boldsymbol{\beta}$, does not address key biological questions
572 about natural selection, such as whether or not fitness optima exist (Schluter, 1988). gppr, on
573 the other hand, provides a method of characterising the major axes of selection, whether they
574 be linear, disruptive, stabilising, or purely directional but curved. Further investigations of the
575 behaviour of gppr-based selection analysis seems warranted, as it is currently unknown what
576 effects details of its application, for example the form of ridge functions, may have on inferences
577 of selection.

578 Reporting of maximum likelihood estimates of selection coefficients, i.e., such as those com-
579 monly reported to date and in table 2, will remain very important. These, and associated
580 information about their statistical uncertainties, are required for meta analysis. To date, failure
581 to report standard errors has severely limited sample sizes in formal meta-analyses of selection
582 (Morrissey & Hadfield, 2012; Siepielski *et al.*, 2013). However, reporting standard errors is only
583 a start. Reporting of sampling variances and covariances, in additional to full reporting of sum-
584 mary statistics, will also be very useful. Sampling covariances could potentially be reported by
585 archiving posterior distributions, or bootstrap samples, of fitted models of fitness functions. An
586 efficient way to report full distributions of sampling variance may be to archive bootstrap dis-
587 tributions of selection coefficients, such as those that are generated automatically by functions
588 in the R package gsg (Morrissey & Sakrejda, 2013).

Conclusion

The availability of an approach to obtain valid inference of selection gradients from arbitrary regression-based inferences of fitness functions renders a large range of techniques available for quantitative inference of natural selection. I have explored a range of these methods, and dissected some key aspects of the behaviour of each. Projection-pursuit regression seems to stand out as a method for characterisation of multivariate selection. Its greater use will facilitate ongoing attempts to implement multivariate quantitative genetic studies on selection in natural populations and experimental systems. Furthermore, identification of major axes of selection, as opposed to major axes of directional or quadratic selection, will bring much more direct biological interpretation to multivariate selection analysis.

4   Acknowledgements

5   Appendix

Denote an arbitrary function relating trait to relative fitness, $w(z)$, and a decomposition of an individual i's trait value, z, into effects of breeding value and environment $z_i = a_i + e_i$ Assume that a and e are independent, $a_i \sim p(a)$ and $e_i \sim q(e)$, such that the variance of phenotype in a population obeys $\sigma_z^2 = \sigma_a^2 + \sigma_e^2$. Assume that $p(a)$ represents a normal probability density function with mean zero and variance σ_a^2, such that $p(a) = \frac{1}{\sigma_a \sqrt{2\pi}} e^{-\frac{a^2}{2\sigma_a^2}}$, and that $q(e)$ is an arbitrary probability density function (not necessarily with mean of zero).

616   The secondary theorem (Robertson, 1966) defines

$$\Delta \bar{z} = \sigma_{a,w}, \tag{A1}$$

617   which by definition is

$$\Delta \bar{z} = E(a \cdot w) - E(a)E(w). \tag{A2}$$

618   The second term in A2 is zero because the $E(a)$ is zero by construction. So, from equation A2,

$$\Delta \bar{z} = \int_{-\infty}^{\infty} a \int_{-\infty}^{\infty} w(a + e)p(a)q(e)deda. \tag{A3}$$

619   The average slope of the relative fitness function, $w'(z)$, given normally distributed breeding
620   values and conditioning on e, can be written $\int_{-\infty}^{\infty} w'(a+e)p(a)da$. $p'(a) = -\frac{a}{\sigma_a^2}p(a)$, so integration
621   by parts gives

$$
\begin{aligned}
\int_{-\infty}^{\infty} w'(a + e)p(a)da &= [w(a + e)p(a)] + \int_{-\infty}^{\infty} w(a + e)\frac{a}{\sigma_a^2}p(a)da \\
&= \int_{-\infty}^{\infty} w(a + e)\frac{a}{\sigma_a^2}p(a)da.
\end{aligned} \tag{A4}
$$

622   The simplification assumes that the relative fitness function is bounded. Applying Fubini's
623   theorem to the double integral and multiplying equation A3 by $\frac{\sigma_a^2}{\sigma_a^2}$, and then substituting using
624   equation A4 gives

$$
\begin{aligned}
\Delta \bar{z} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma_a^2 w(a + e)\frac{a}{\sigma_a^2}p(a)q(e)dade, \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma_a^2 w'(a + e)p(a)q(e)dade, \\
&= \sigma_a^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w'(a + e)p(a)q(e)dade.
\end{aligned} \tag{A5}
$$

625   So, evolution of the mean phenotype is given by the variance of normally distributed breeding
626   values, times the average slope of the relative fitness function integrated over the distribution of
627   phenotype, regardless of the distribution of environmental effects on phenotype.

628   # 6   Data accessibility

629   R functions and example datasets are included in an update to the R package GSG.

630   # References

631   Bell, G. (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environ-

632   ments. *Philosophical Transactions of the Royal Society, Series B*, **365**, 87–97.

633 Blows, M.W. (2007) A tale of two matrices: multivariate approaches in evolutionary biology.
634   *Journal of Evolutionary Biology*, **20**, 1–8.

635 Bolnick, D.I. & Lau, O.L. (2008) Predictable patterns of disruptive selection in stickleback in
636   postglacial lakes. *The American Naturalist*, **172**, 1–11.

637 Carlson, S.M. & Quinn, T.P. (2007) Ten years of varying lake level and selection on size-at-
638   maturity in sockeye salmon. *Ecology*, **88**, 2620–2629.

639 Clutton-Brock, T.H. & Pemberton, J.M. (2004) *Soay sheep dynamics and selection in an island*
640   *population.* Cambridge University Press, Cambridge.

641 Clutton-Brock, T.H., Price, O.F., Albon, S.D. & Jewell, P.A. (1992) Early development and
642   population fluctuations in soay sheep. *Journal of Animal Ecology*, **61**, 381–396.

643 Coltman, D.W., Smith, J.A., Bancroft, D.R., Pilkington, J., MacColl, A.D.C., Clutton-Brock,
644   T.H. & Pemberton, J.M. (1999) Density-dependent variation in lifetime breeding success and
645   natural sexual selection in soay rams. *The American Naturalist*, **154**, 730–746.

646 Dickerson, G.E. (1955) Genetic slippage in response to selection for multiple objectives. *Cold*
647   *Spring Harbor Symposia on Quantitative Biology*, **20**, 213–224.

648 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gru-
649   ber, B., Lafourcade, B., ao, P.J.L., Münkemüller, T., McClean, C., Osborne, P.E., Reineking,
650   B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2013) Collinearity: a review of
651   methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**,
652   27–46.

653 Endler, J.A. (1986) *Natural selection in the wild.* Princeton University Press.

654 Engen, S. & Saether, B.E. (2014) Evolution in fluctuating environments: decomposing selection
655   into additive components of the Robertson-Price equation. *Evolution*, **68**, 854–865.

656 Engen, S., Saether, B.E., Kvalnes, T. & Jensen, H. (2012) Estimating fluctuating selection in
657   age-structured populations. *Journal of Evolutionary Biology*, **25**, 1487–1499.

658 Friedman, J., Hastie, T. & Tibshirani, R. (2008) Regularization paths for generalized linear

659   models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.

660 Friedman, J.H. & Stuetzle, W. (1981) Projection pursuit regression. *Journal of the American*

661   *Statistical Association*, **76**, 817–823.

662 Geyer, C.J., Wagenius, S. & Shaw, R.G. (2007) Aster models for life history analysis. *Biometrika*,

663   **94**, 415–426.

664 Grant, P.R. & Grant, B.R. (2002) Unpredictable evolution in a 30-year study of Darwin's finches.

665   *Science*, **296**, 707–711.

666 Grether, G.F. (1996) Sexual selection and survival selection on wing coloration and body size in

667   the rubyspot damselfly *Hataerina americana*. *Evolution*, **50**, 1939–1948.

668 Hansen, T.F. & Houle, D. (2008) Measuring and comparing evolvability and constraint in mul-

669   tivariate characters. *Journal of Evolutionary Biology*, **21**, 1201–1219.

670 Hereford, J., Hansen, T.F. & Houle, D. (2004) Comparing strengths of directional selection: how

671   strong is strong? *Evolution*, **58**, 2133–2143.

672 Johnston, S.E., Gratten, J., Berenos, C., Pilkington, J.G., Clutton-Brock, T.H., Pemberton,

673   J.M. & Slate, J. (2013) Life history trade-offs at a single locus maintain sexually selected

674   genetic variation. *Nature*, **502**, 93–95.

675 Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Vignieri, C., Berrigan, D., Hill, E., Hoang, A.,

676   Gilbert, P. & Beerli, P. (2001) The strength of phenotypic selection in natural populations.

677   *The American Naturalist*, **157**, 245–261.

678 Lande, R. & Arnold, S.J. (1983) The measurement of selection on correlated characters. *Evolu-*

679   *tion*, **37**, 1210–1226.

680 Lockhart, R., Taylor, J., Tibshirani, R.J. & Tibshirani, R. (2013) A significance test for the

681   lasso. *available at* `http://arxivorg`, **arXiv:1301.7161v2**.

682 MacColl, A.D.C. (2011) The ecological causes of evolution. *Trends in Ecology and Evolution*,

683   **26**, 514–522.

684 Milner, J.M., Albon, S.D., Illius, A.W., Pemberton, J.M. & Clutton-Brock, T.H. (1999) Repeated
685   selection of morphometric traits in the Soay sheep of St Kilda. *Journal of Animal Ecology*, **68**,
686   472–488.

687 Mitchell-Olds, T. & Shaw, R.G. (1987) Regression analysis of natural selection: statistical infer-
688   ence and biological interpretation. *Evolution*, **41**, 1149–1161.

689 Morrissey, M.B. (2014) Selection and evolution of causally covarying traits. *Evolution*, **68**, 1748–
690   1761.

691 Morrissey, M.B. (in review) Severe biases in synthetic meta-analyses that fail to account for the
692   observation process. *The American Naturalist*.

693 Morrissey, M.B. & Hadfield, J.D. (2012) Directional selection in temporally replicated studies is
694   remarkably constant. *Evolution*, **66**, 435–442.

695 Morrissey, M.B., Parker, D.J., Korsten, P., Pemberton, J.M., Kruuk, L.E.B. & Wilson, A.J.
696   (2012a) The prediction of adaptive evolution: empirical application of the secondary theorem
697   of selection and comparison to the breeder's equation. *Evolution*, **currently online early**.

698 Morrissey, M.B. & Sakrejda, K. (2013) Unification of regression-based approaches to the analysis
699   of natural selection. *Evolution*, **67**, 2094–2100.

700 Morrissey, M.B., Walling, C.A., Wilson, A.J., Pemberton, J.M., Clutton-Brock, T.H. & Kruuk,
701   L.E.B. (2012b) Genetic anlaysis of life history constraint and evolution in a wild ungulate
702   popualtion. *American Naturalist*, **179**, E97–E114.

703 Philips, P.C. & Arnold, S.J. (1989) Visualizing multivariate selection. *Evolution*, **43**, 1209–1222.

704 Plummer, M. (2010) *JAGS version 2.0 Manual*. International Agency for Research on Cancer.

705 Poissant, J., Wilson, A.J., Festa-Bianchet, M., Hogg, J.T. & Coltman, D.W. (2008) Quantitative
706   genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings*
707   *of the Royal Society, B Series*, **275**, 623–628.

708 Reynolds, R.J., Childers, D.K. & Pajewski, N.M. (2010) The distribution and hypothesis testing
709   of eigenvalues from the canonical analysis of the gamma matrix of quadratic and correlational
710   selection gradients. *Evolution*, **64**, 1076–1085.

711 Rice, W.R. (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

712 Robertson, A. (1966) A mathematical model of the culling process in dairy cattle. *Animal*
713   *Production*, **8**, 95–108.

714 Robinson, M.R., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Kruuk, L.E.B.
715   (2006) Live fast, die young: a balance of natural, sexual and antagonistic selection maintains
716   phenotypic polymorphism of weaponry in male and female soay sheep. *Evolution*, **60**, 2168–
717   2181.

718 Roff, D.A. & Fairbairn, D.J. (2007) The evolution of trade-offs: where are we? *Journal of*
719   *Evolutionary Biology*, **20**, 433–447.

720 Schluter, D. (1988) Estimating the form of natural selection on a quantitative trait. *Evolution*,
721   **42**, 849–861.

722 Schluter, D. & Nychka, D. (1994) Exploring fitness surfaces. *The American Naturalist*, **143**,
723   597–616.

724 Schluter, D. & Smith, J.N.M. (1986) Natural selection on beak and body size in the song sparrow.
725   *Evolution*, **40**, 221–231.

726 Shaw, R.G. & Geyer, C.J. (2010) Inferring fitness landscapes. *Evolution*, **64**, 2510–2520.

727 Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. (2009) It's about time: the temporal dynamics
728   of phenotypic selection in the wild. *Ecology Letters*, **12**, 1261–1276.

729 Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., DiBattista, J.D. & Carlson,
730   S.M. (2013) The spatial patterns of directional phenotypic selection. *Ecology Letters*.

731 Sober, E. (1984) *The Nature of Selection*. University of Chicago Press, Chicago.

732 Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal*
733   *Statistical Society, Series B*, **58**, 267–288.

734 Tikhonov, A.N. & Arsenin, V.Y. (1977) *Solution of ill-posed problems.* Winston and sons,
735   Washington.

736 Walsh, B. & Blows, M.W. (2009) Abundant genetic variation + strong selection = multivariate
737   genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution,*
738   *and Systematics*, **40**, 41–59.

739 Wilson, K., Grenfell, B.T., Pilkington, J.G., Boyd, H.E.G. & Gulland, F.M.D. (2004) *Parasites*
740   *and their impact, in Soay sheep: dynamics and selection in an island populaiton, T. H. Clutton-*
741   *Brock and J. M. Pemberton, eds.* Cambridge University Press.

742 Wood, S.N. (2006) *Generalized Additive Models: An Introduction with R.* Chapman and
743   Hall/CRC.

744 Zou, H. & Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of*
745   *the Royal Statistical Society, Series B*, **67**, 301–320.

Table 1: The distribution of mass, hind leg length, horn length, and log ked number in Soay sheep lambs. (a) Correlation matrices (all associated analyses are conducted on unit variance-standardised data for both sexes in crash and non-crash years), and (b) spectral decomposition of the correlation matrices. Units are kg for mass, and mm for leg and horn lengths.

(a) correlation, means and variances	mass	leg length	horn length		mean (SE)	variance (SE)
(a*i*) males, non-crash years						
mass					13.85 (0.11)	7.79 (0.44)
leg length	0.84				162.00 (0.40)	98.99 (5.57)
horn length	0.573	0.421			126.72 (1.53)	1484.8 (83.5)
(log) ked number	-0.025	0.005	0.062		1.269 (0.028)	0.515 (0.029)
(a*ii*) females, non-crash years						
mass					12.35 (0.15)	4.79 (0.46)
leg length	0.861				157.63 (0.64)	86.64 (8.42)
horn length	0.472	0.425			83.06 (1.55)	511.2 (49.6)
(log) ked number	-0.105	-0.056	-0.052		0.978 (0.046)	0.460 (0.45)
(a*iii*) males, crash years						
mass					13.54 (0.19)	9.95 (0.84)
leg length	0.864				161.21 (0.64)	117.47 (9.93)
horn length	0.594	0.507			125.94 (2.35)	1555.7 (131.5)
(log) ked number	0.006	0.005	-0.004		1.21 (0.045)	0.563 (0.048)
(a*iv*) females, crash years						
mass					12.56 (0.24)	6.60 (0.87)
leg length	0.856				157.08 (0.85)	83.75 (11.00)
horn length	0.509	0.454			83.40 (2.11)	521.5 (68.48)
(log) ked number	-0.329	-0.286	-0.154		1.124 (0.074)	0.637 (0.084)

(b) major axes of correlation matrices					
	eigenvalue	eigenvector loadings			
		mass	leg length	horn length	ked number
(b*i*) males, non-crash years					
term 1	2.224	0.633	0.594	0.496	0.017
term 2	1.009	0.070	0.051	-0.115	-0.990
term 3	0.621	0.182	0.499	-0.836	0.136
term 4	0.146	0.749	-0.629	-0.205	0.044
(b*iii*) females, non-crash years					
term 1	2.211	0.628	0.615	0.465	-0.103
term 2	0.988	0.031	0.085	0.065	0.994
term 3	0.665	0.292	0.371	-0.881	0.017
term 4	0.136	0.721	-0.690	-0.051	0.040
(b*iii*) males, crash years					
term 1	2.322	0.619	0.599	0.508	0.004
term 2	1.000	0.004	0.004	-0.016	1.000
term 3	0.548	0.267	0.447	-0.853	-0.016
term 4	0.129	0.738	-0.664	-0.117	-0.002
(b*iv*) females, crash years					
term 1	2.383	0.601	0.584	0.447	-0.314
term 2	0.871	0.082	0.107	0.389	0.911
term 3	0.606	0.314	0.434	-0.802	0.264
term 4	0.140	0.731	-0.678	-0.067	0.042

Table 2: Variance standardised selection differentials and gradients for summer lamb traits in male and female Soay sheep in crash and non-crash years.

(a) differentials

| | Non-crash years | | | | | | Crash years | | | | | |
| | males | | | females | | | males | | | females | | |
	estimate	SE	p	estimate	SE	p	estimate	SE	p	estimate	SE	p
S, mass	**0.133**	**0.027**	**< 0.001**	**0.179**	**0.037**	**< 0.001**	**0.293**	**0.076**	**< 0.001**	**0.396**	**0.094**	**< 0.001**
S, leg length	**0.113**	**0.027**	**< 0.001**	**0.163**	**0.039**	**< 0.001**	**0.253**	**0.075**	**0.002**	**0.38**	**0.085**	**< 0.001**
S, horn length	0.011	0.027	0.666	0.015	0.030	0.614	**0.222**	**0.078**	**0.008**	**0.364**	**0.089**	**< 0.001**
S, ked number	0.004	0.027	0.890	-0.047	0.035	0.178	0.087	0.077	0.272	-0.100	0.099	0.328
C, mass	0.003	0.035	0.948	-0.053	0.061	0.466	0.030	0.121	0.780	-0.050	0.138	0.700
C, leg length	-0.070	0.044	0.094	-0.098	0.065	0.136	-0.062	0.105	0.572	-0.229	0.109	0.060
C, horn length	0.003	0.032	0.934	0.093	0.045	0.052	-0.042	0.088	0.612	-0.197	0.121	0.114
C, ked number	0.023	0.033	0.436	-0.001	0.036	0.962	-0.052	0.102	0.580	-0.236	0.126	0.106
C, mass - leg length	-0.029	0.034	0.378	-0.061	0.063	0.388	-0.033	0.101	0.768	-0.118	0.115	0.292
C, mass - horn length	-0.011	0.030	0.698	-0.002	0.043	0.956	-0.009	0.094	0.922	0.057	0.128	0.682
C, mass - ked number	0.012	0.025	0.632	-0.030	0.038	0.442	-0.044	0.079	0.576	0.058	0.084	0.504
C, leg length - horn length	-0.032	0.029	0.258	-0.024	0.047	0.592	-0.057	0.083	0.516	-0.017	0.119	0.838
C, leg length - ked number	0.016	0.026	0.552	-0.046	0.041	0.254	-0.053	0.076	0.490	0.150	0.086	0.098
C, horn length - ked number	0.064	0.028	0.016	-0.027	0.028	0.330	-0.014	0.076	0.852	-0.062	0.087	0.518

(b) gradients

| | Non-crash years | | | | | | Crash years | | | | | |
| | males | | | females | | | males | | | females | | |
	estimate	SE	p	estimate	SE	p	estimate	SE	p	estimate	SE	p
β, mass	**0.217**	**0.056**	**< 0.001**	**0.168**	**0.067**	**0.013**	0.256	0.181	0.159	0.269	0.22	0.225
β, leg length	-0.036	0.052	0.494	0.062	0.066	0.348	-0.002	0.171	0.989	0.187	0.207	0.368
β, horn length	**-0.108**	**0.035**	**0.002**	**-0.109**	**0.039**	**0.006**	0.064	0.109	0.56	0.108	0.149	0.473
β, ked number	0.011	0.027	0.679	-0.048	0.034	0.157	0.082	0.081	0.314	0.08	0.109	0.464
γ, mass	-0.013	0.009	0.147	0.000	0.008	0.980	0.003	0.014	0.843	-0.015	0.031	0.641
γ, leg length	-0.006	0.003	0.092	-0.003	0.009	0.771	-0.002	0.014	0.871	0.002	0.014	0.883
γ, horn length	-0.009	0.008	0.259	0.000	0.001	0.955	-0.011	0.025	0.655	-0.016	0.012	0.184
γ, ked number	0.003	0.006	0.563	0.007	0.006	0.214	-0.001	0.014	0.971	-0.002	0.007	0.779
γ, mass - leg length	0.058	0.051	0.258	-0.026	0.054	0.629	0.011	0.152	0.945	0.194	0.177	0.277
γ, mass - horn length	0.103	0.059	0.081	0.101	0.084	0.234	-0.029	0.202	0.884	0.053	0.268	0.843
γ, mass - ked number	-0.061	0.055	0.273	0.100	0.066	0.131	0.029	0.182	0.875	0.036	0.235	0.879
γ, leg length - horn length	-0.062	0.055	0.256	-0.036	0.089	0.688	0.045	0.185	0.808	-0.036	0.236	0.880
γ, leg length - ked number	0.050	0.051	0.324	-0.102	0.064	0.112	-0.055	0.171	0.75	0.112	0.203	0.583
γ, horn length - ked number	0.079	0.032	0.015	-0.006	0.046	0.900	0.023	0.109	0.833	-0.073	0.140	0.603

Table 3: Quadratic selection gradients of the major axes of estimated γ; obtained as the eigenvalues of the the diagonalised estimated γ matrix for each sex and environmental condition. p-values are permutation-based two-tailed values following Reynolds *et al.* (2010). The first two eigenvectors for each analysis are depicted graphically in figure 1.

axis of γ	1	p	2	p	3	p	4	p
non-crash, males	-0.222	0.497	0.070	0.845	0.024	0.926	-0.021	0.436
non-crash, females	-0.475	0.224	-0.199	0.23	0.159	0.068	-0.017	0.627
crash, males	0.37	0.935	-0.137	0.978	-0.074	0.925	0.016	0.867
crash, females	-1.874	0.069	-0.373	0.686	0.089	0.947	-0.069	0.559

Table 4: Variance standardised selection gradients for summer lamb traits in male and female Soay sheep in crash and non-crash years using regularised regression and projection-pursuit regression. Zero values with no decimal places shown are quadratic gradients associated with traits that have both associated directional and quadratic coefficient estimates shrunken to zero.

(a) Non-crash years

	males						females					
	lasso	ridge	elastic net	gppr	gppr SE	gppr p	lasso	ridge	elastic net	gppr	gppr SE	gppr p
β, mass	0.145	0.094	0.121	0.214	0.057	0	0.153	0.117	0.126	0.194	0.063	0.004
β, leg length	0	0.039	0.006	-0.023	0.048	0.596	0.022	0.079	0.052	0.038	0.057	0.488
β, horn length	-0.051	-0.046	-0.036	-0.105	0.034	0	-0.054	-0.086	-0.062	-0.121	0.044	0
β, ked number	-0.001	0.007	-0.001	0.014	0.026	0.628	-0.010	-0.030	-0.019	-0.032	0.032	0.316
γ, mass	-0.024	0.019	-0.017	-0.045	0.027	0.01	-0.067	-0.036	-0.040	-0.081	0.054	0.018
γ, leg length	-0.003	-0.035	-0.002	0	0.004	0.674	-0.002	-0.078	-0.009	0	0.015	0.668
γ, horn length	-0.005	-0.006	-0.003	-0.01	0.009	0.022	-0.007	-0.002	-0.008	-0.027	0.025	0.018
γ, ked number	-0.002	0.019	-0.002	0	0.001	0.916	-0.001	-0.002	-0.003	0.001	0.005	0.916
γ, mass - leg length	0	0.004	-0.001	0.005	0.013	0.594	-0.010	0.014	-0.019	-0.020	0.025	0.492
γ, mass - horn length	0.009	0.015	0.005	0.023	0.013	0.01	0.035	0.058	0.038	0.050	0.029	0.004
γ, mass - ked number	0	-0.018	0	-0.003	0.006	0.596	0.003	-0.004	0.008	0.008	0.017	0.492
γ, leg length - horn length	-0.001	-0.015	-0.001	-0.003	0.007	0.626	0.006	0.036	0.012	0.013	0.015	0.326
γ, leg length - ked number	0	0.013	0	0	0.002	0.972	-0.011	-0.048	-0.021	0	0.006	0.888
γ, horn length - ked number	0.043	0.048	0.038	0.001	0.003	0.634	-0.002	-0.014	-0.004	-0.01	0.011	0.316

(b) Crash years

	males						females					
	lasso	ridge	elastic net	gppr	gppr SE	gppr p	lasso	ridge	elastic net	gppr	gppr SE	gppr p
β, mass	0.182	0.060	0.154	0.233	0.169	0.166	0.133	0.137	0.126	0.176	0.364	0.340
β, leg length	0	0.046	0.021	0.019	0.158	0.836	0.083	0.118	0.102	0.158	0.370	0.424
β, horn length	0.009	0.042	0.036	0.077	0.095	0.43	0.133	0.108	0.13	0.216	0.327	0.038
β, ked number	0	0.022	0.006	0.088	0.077	0.232	0	0.008	0	0.041	0.329	0.674
γ, mass	0.014	0.010	0.010	0.016	0.060	0.368	0.002	-0.004	0.001	-0.006	12.657	0.496
γ, leg length	0	-0.003	0	0	0.044	0.684	0.001	-0.023	0.001	-0.006	12.657	0.242
γ, horn length	0	-0.016	0.001	0	0.013	0.812	0.002	-0.084	-0.009	-0.007	12.657	0.430
γ, ked number	0	-0.018	0	0.001	0.012	0.69	0	-0.059	-0.001	-0.002	12.654	0.170
γ, mass - leg length	0	0	0.001	0.002	0.048	0.978	0.001	0.014	0.001	0.002	12.651	0.812
γ, mass - horn length	0.001	0.002	0.002	0.007	0.020	0.53	0.002	0.029	0.002	0.002	0.220	0.572
γ, mass - ked number	0	-0.004	0	0.001	0.017	0.774	0.001	0.013	0.002	0.002	12.65	0.898
γ, leg length - horn length	0	-0.005	0	0.009	0.016	0.276	0	0.012	0	0.001	0.111	0.598
γ, leg length - ked number	0	-0.008	0	0.001	0.009	0.758	0	0.041	0.005	0	0.111	0.614
γ, horn length - ked number	0	0.001	0	0.003	0.010	0.416	0	-0.04	0	0.001	0.110	0.602

Table 5: Selection gradients (non-standardised) of major axes of (standardised) **P** in male and female Soay sheep in crash and non-crash years.

| | Non-crash years | | | | | | Crash years | | | | | |
| | males | | | females | | | males | | | females | | |
	estimate	SE	p	estimate	SE	p	estimate	SE	p	estimate	SE	p
β, axis 1	**0.075**	**0.033**	**0.020**	**0.101**	**0.053**	**0.048**	**0.194**	**0.071**	**0.008**	**0.264**	**0.097**	**0.006**
β, axis 2	-0.006	0.025	0.818	-0.03	0.033	0.434	0.084	0.079	0.294	0.161	0.1	0.124
γ, axis 1	-0.004	0.017	0.862	-0.005	0.016	0.686	0.01	0.045	0.868	-0.005	0.047	0.756
γ, axis 2	0.04	0.038	0.312	-0.022	0.062	0.758	-0.068	0.112	0.562	-0.207	0.153	0.154
γ, axis 1 - axis 2	0.026	0.016	0.114	0.015	0.019	0.312	-0.011	0.05	0.792	0.059	0.057	0.272

Table 6: Major multivariate axes of selection. The variance of absolute fitness associated with the two main axes of generalised projection-pursuit regression-based inferences of fitness functions. The shape of the first axes of selection, and associated trait loadings, are are depicted graphically in figure 3.

	$\sigma^2_W|t_1$	p	$\sigma^2_W|t_2 - \sigma^2_W|t_2$	p
non-crash, males	**0.012**	**0.000**	**0.0027**	**0.026**
non-crash, females	**0.029**	**0.001**	0*	0.964
crash, males	**0.013**	**0.005**	0*	0.959
crash, females	**0.046**	**0.002**	0.001	0.625

*the estimates reported as 0 were estimated as approximately -0.001. I attribute the slightly lower predicted variance of expected fitness in the two term model to slight imperfections in model fit.
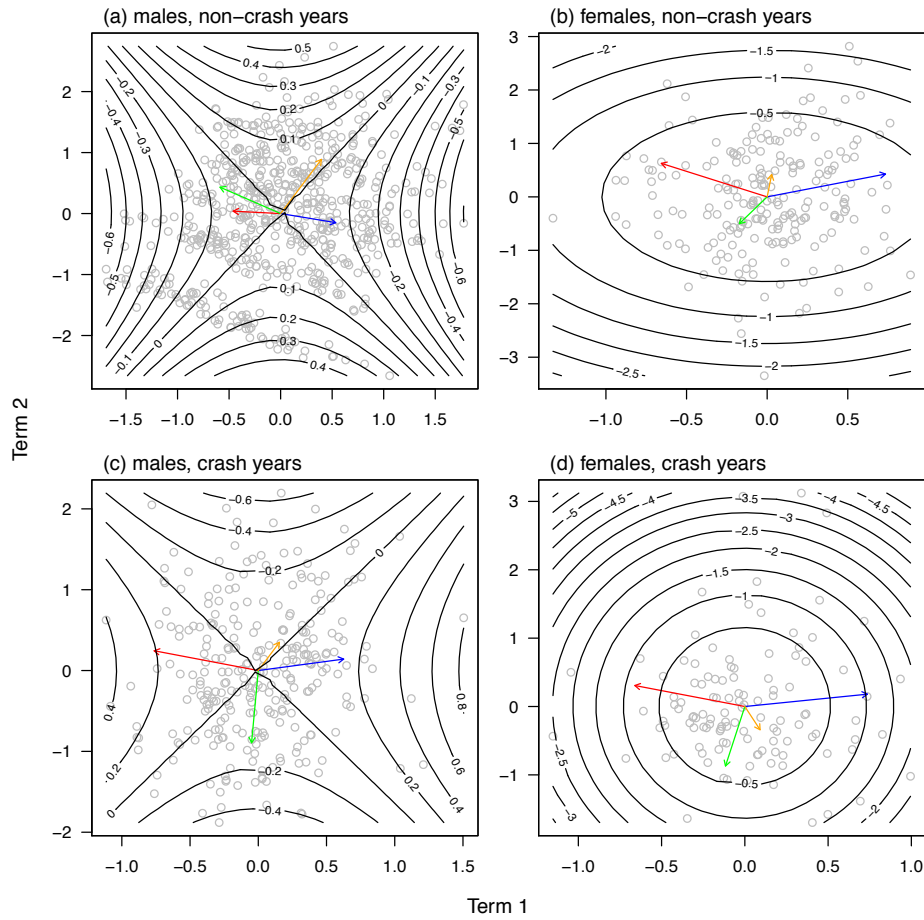
Figure 1: Major axes of estimated γ. The estimated values of γ are given in table 2 and permutation test-based inferences of the statistical significance of quadratic selection of these axes are given in table 6. The coloured arrows represent the loadings of the traits (blue - mass, red - leg length, green - horn length, and orange - log ked count) onto the space defined by the first two eigenvectors of γ. Selection of no axes of γ in either sex or environmental condition are statistically significant.
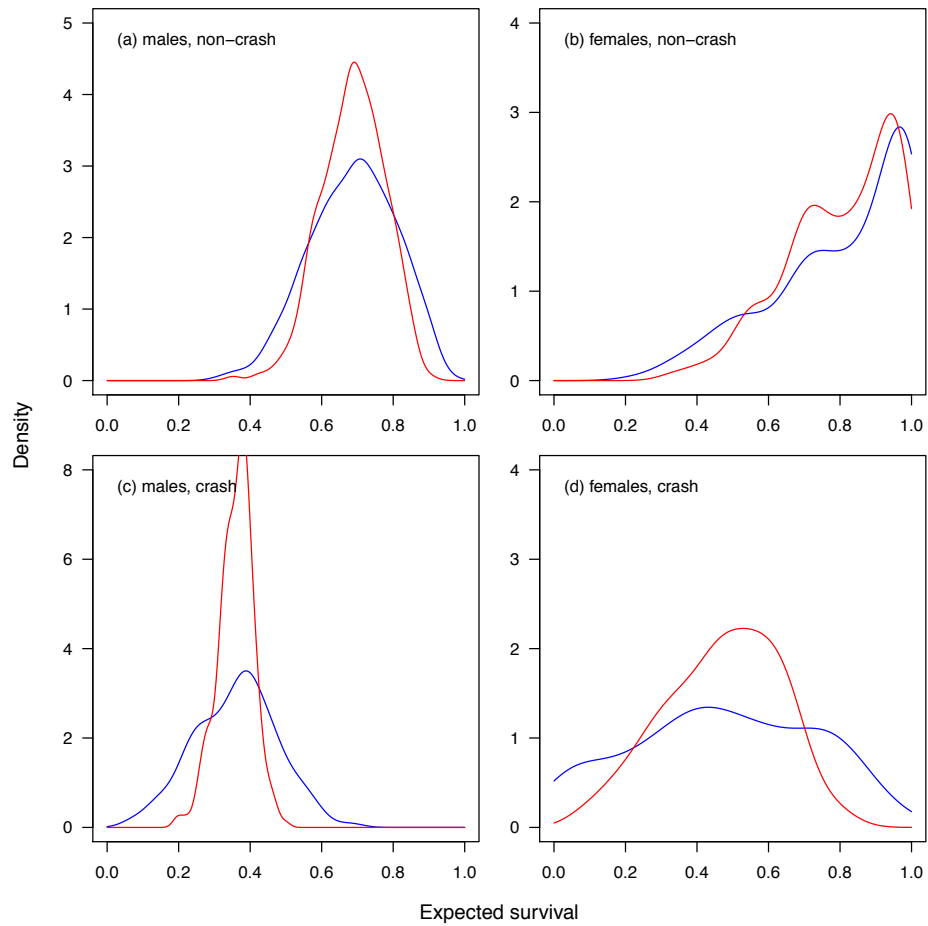
Figure 2: Distributions of expected fitness from the generalised-linear model-based estimates of selection gradients (blue), and from ridge regression-based estimates of selection gradients (red).
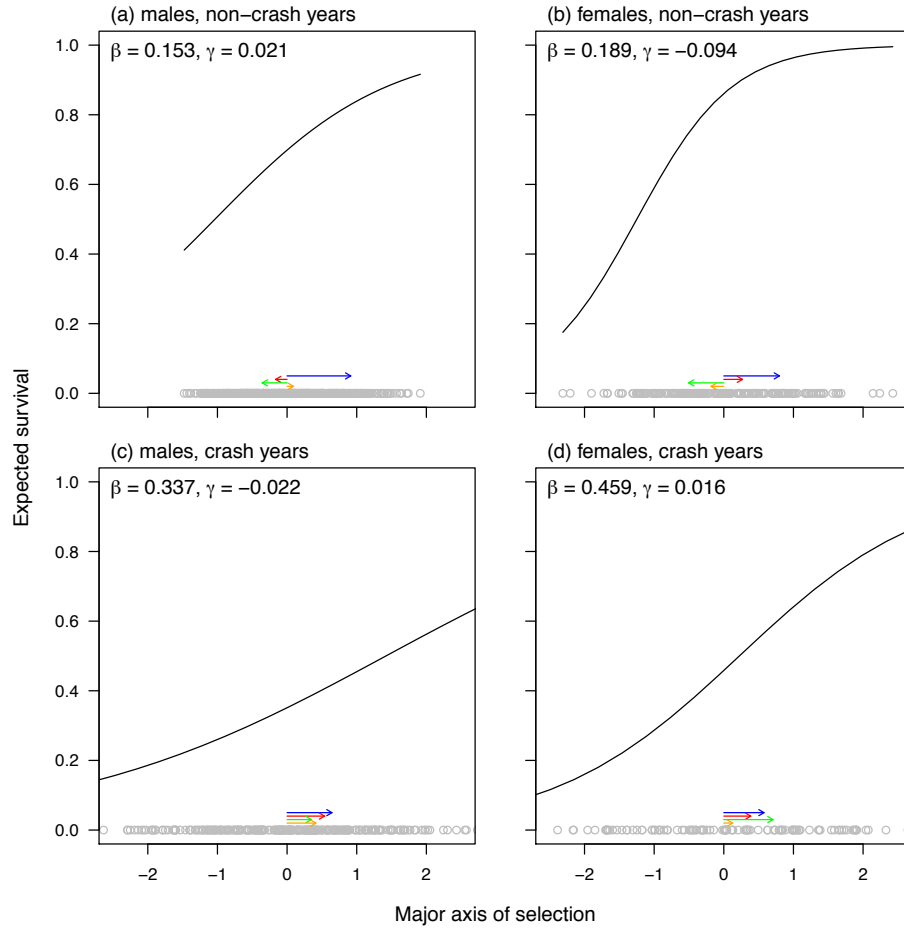
Figure 3: Generalised projection-pursuit regression-based fitness functions for the major axes of selection of Soay sheep lambs. Grey points indicate the values of the four traits transformed to the two major axes of the projection-pursuit regression for each of males (a) and (c), and females (b) and (d), in non-crash (a) and (b), and crash (c) and (d) years, respectively. Coloured arrows describe the rotation of the four traits onto the two major axes: blue - mass, red - leg length, green - horn length, and orange - log ked count. Contours show expected absolute fitness. Values of β and γ reported on each of the plots are the unit variance standardised directional and quadratic selection gradients, and the x-axis is also plotted for unit variance scaled factors. They are obtained by projecting the phenotype onto the major axis of selection, as inferred by gppr, and then fitting a univariate cubic spline to the rotated phenotype, predicting individual survival, and the using the GSG function `gam.gradients()` to obtain the selection gradients.
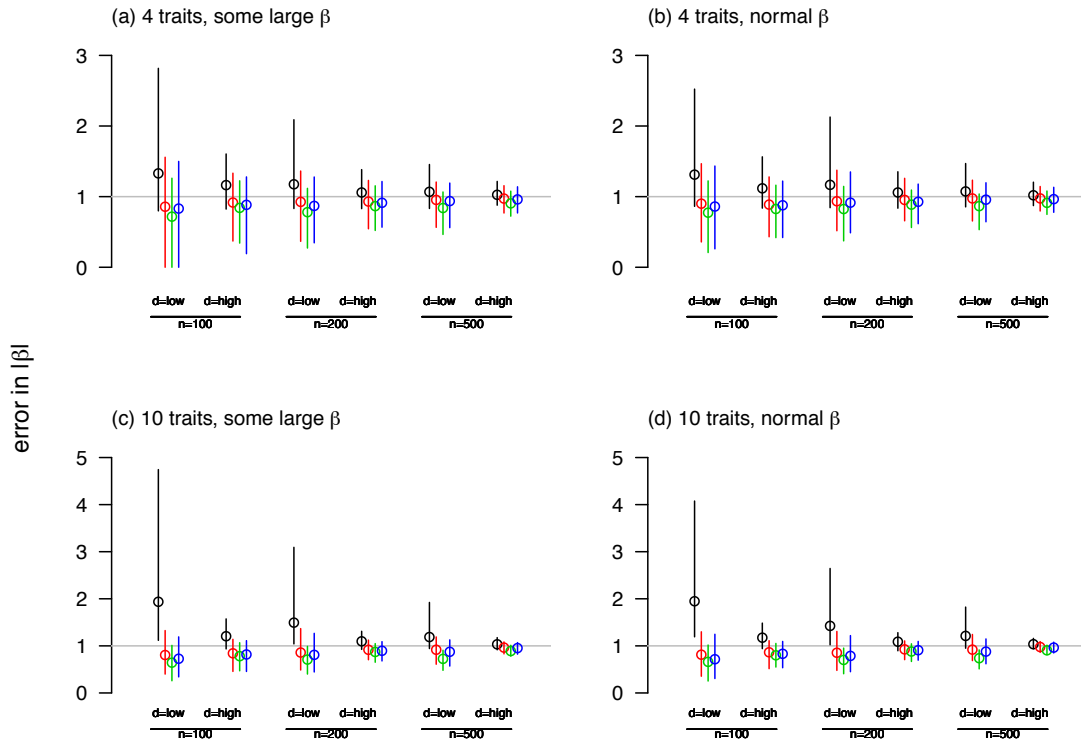
Figure 4: Proportional error in the length of estimated selection gradient vectors. Points represent means, and lines represent 80% quantile ranges of the difference between estimated and true values of $||\boldsymbol{\beta}||$. Mean values greater than one represent positive bias, and values below one demonstrate negative bias. Colours represent: black - non-regularised GLM regression, red - ridge regression, green - lasso, and blue - elastic net. "d = high or low" indicates covariance matrices with high and low dimensionality, "some large β" in parts (a) and (c) indicate β in those simulations are drawn from a t-distribution, rather than a normal distribution; further details on simulation scenarios are given in the text.
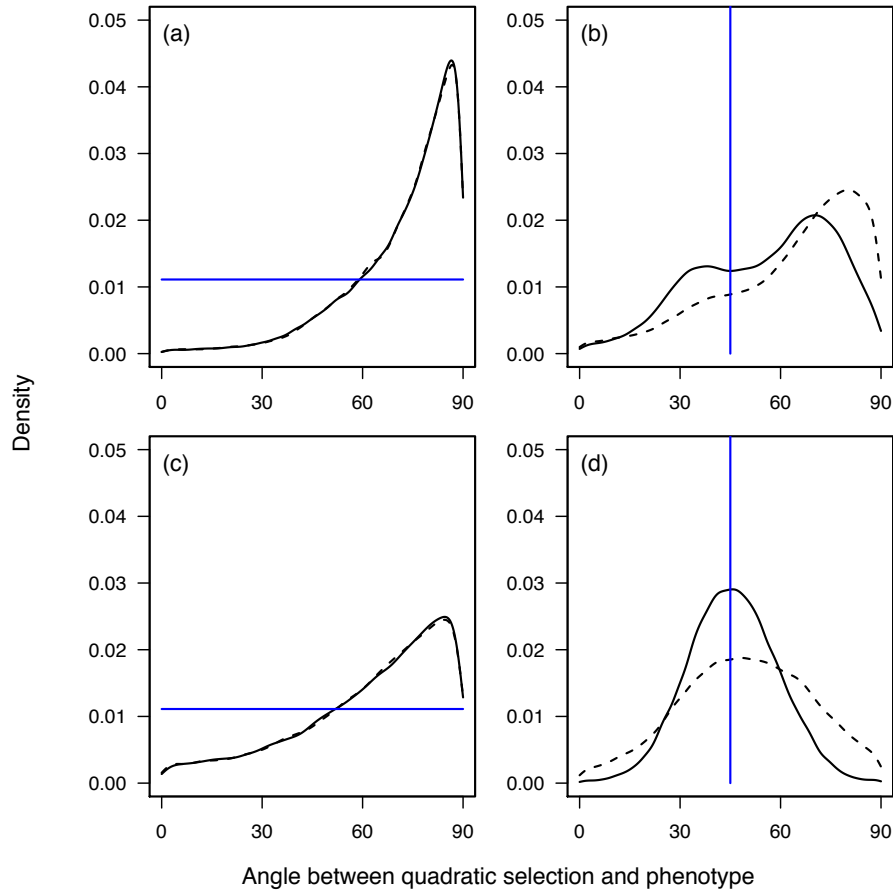
Figure 5: Simulation results showing bias in the direction of major axes of estimated γ matrices. All plots show the distribution of angles between the major axis of phenotype, and the major axis of the estimated γ matrix. Left plots (a and c): no selection, right plots (b and d) quadratic selection of one trait; top plots: phenotypic correlations of 0.8, and bottom plots: $r_P = 0.5$. Solid lines show simulations for sample size of 500, and dotted lines show n = 250. Each simulation scenario was repeated 50000 times. In (a and c), the blue lines show the distribution of an hypothetical unbiased estimator. In (b and d), the blue lines show the true value around which an unbiased estimator would be distributed.