1   # Variation in reaction norms: statistical considerations and biological

2   # interpretation

3   Michael B. Morrissey[1,3], and Maartje Liefting[2]

4   June 15, 2016

5   [1]School of Biology, University of St Andrews
6   [2]Department of Animal Ecology, VU University Amsterdam
7

   [3]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

8 **Abstract**

9 Analysis of reaction norms, the functions by which the phenotype produced by a given geno-
10 type depends on the environment, is critical to studying many aspects of phenotypic evo-
11 lution. Different techniques are available for quantifying different aspects of reaction norm
12 variation. We examine what biological inferences can be drawn from some of the more readily-
13 applicable analyses for studying reaction norms. We adopt a strongly biologically-motivated
14 view, but draw on statistical theory to highlight strengths and drawbacks of different tech-
15 niques. In particular, consideration of some formal statistical theory leads to revision of
16 some recently, and forcefully, advocated opinions on reaction norm analysis. We clarify what
17 simple analysis of the slope between mean phenotype in two environments can tell us about
18 reaction norms, explore the conditions under which polynomial regression can provide ro-
19 bust inferences about reaction norm shape, and explore how different existing approaches
20 may be used to draw inferences about variation in reaction norm shape. We show how mixed
21 model-based approaches can provide more robust inferences than more commonly-used multi-
22 step statistical approaches, and derive new metrics of the relative importance of variation in
23 reaction norm intercepts, slopes, and curvatures.

24 Introduction

25 Characterising the functions describing the dependence upon the environment of phenotypes
26 generated by specific genotypes is critical to understanding many aspects of evolution. These
27 functions, termed reaction norms (Schmalhausen, 1949; Woltereck, 1909), are therefore the
28 subject of a great deal of interest from evolutionary biologists (Gupta and Lewontin, 1982;
29 Scheiner, 1993; West-Eberhard, 2003). For example, characterisations of reaction norms can
30 be important for understanding how populations will respond to changing environments,
31 and so the extent to which non-evolutionary plastic responses and adaptive evolutionary
32 change can allow populations to persist (Chevin et al., 2010; Ghalambor et al., 2007). In
33 microevolutionary studies, we may often be interested both in the mean reaction norm of
34 populations, and also in variation in reaction norms within populations (Nussey et al., 2005).
35 Assessment of variation in reaction norms can in principle inform us of how traits will evolve
36 in response to selection across a range of environments (Kirkpatrick et al., 1990; Scheiner
37 and Callahan, 1999).

38 The true shapes of reaction norms are potentially complex, and any empirical analysis
39 will require a model of reaction norms (DeWitt and Scheiner, 2004; Gavrilets and Scheiner,
40 1993). Two general principles of models will hold true for the analysis of reaction norms.
41 First, models of reaction norms will typically be simpler than the true (unknown) functions
42 themselves. This simplification is not a weakness of model-based approaches, but in fact is
43 key to generating tractable inferences. Second, simple models may also have properties that
44 do not reflect, or only poorly reflect, some properties of true reaction norms. These two
45 general principles will invariably apply both to model-based inferences of specific reaction
46 norms (e.g., the average response of a genotype or population to an environmental variable),
47 and to inferences about variation in reaction norms (e.g., inferences of the amount of variation
48 in say, the steepness of reaction norms among different genotypes in a population).

49 The primary goal of this paper is to examine how some of the most readily-applicable
50 statistical models of reaction norms can be used to make robust inferences about properties

51 of reaction norms. As examples of such reaction norm properties, we consider inference both

52 of properties of reaction norms of individual genotypes (or other genetic groupings such as

53 populations; e.g., focusing on their slopes, or the locations of their maxima), and properties of

54 families of reaction norms (e.g., variation in slopes, or variation in the locations of maxima).

55 The primary focus is on biological inference, but we draw extensively on the statistical

56 theory underlying different potential analytical approaches to studying reaction norms. In

57 some cases, we expand basic theory about regression analysis to yield new insights about

58 how specific, biologically-motivated, regression analyses may behave. We discuss biological

59 inference of properties of reaction norms in general, but we also specifically focus on on some

60 recent claims that have been made about the efficacy of different approaches.

61 Polynomial regression, and especially quadratic regression, is potentially very useful for

62 characterising reaction norms, and several authors have investigated theoretical and empir-

63 ical properties of reaction norms using such functions as theoretical and statistical models

64 (e.g., Delpuech et al. 1995; Gavrilets and Scheiner 1993). Two recent very firm claims about

65 analysis of reaction norms with polynomial functions are: (1) that the slope of a line con-

66 necting mean phenotype in two environments is generally misleading about the form of a

67 reaction norm (Rocha and Klaczko, 2012); and (2) that quantities derivable from polynomial

68 regressions, such as the slope at any point, or measures of overall curvature, provide robust

69 inference of reaction norms (Rocha and Klaczko, 2014). We show analytically, and with

70 numerical examples, that neither of these assertions is generally true. Nonetheless, we agree

71 that polynomial regression, perhaps especially quadratic regression, may be very useful for

72 biological studies of reaction norms. However, polynomial regression will be most useful if

73 applied with a somewhat more nuanced understanding of its strengths and limitations.

74 We also contrast two approaches to characterising variation in reaction norms. By "char-

75 acterising variation", we refer to situations where we are not necessarily interested in specific

76 reaction norms, nor in comparisons of properties of two or few specific reaction norms, but

77 rather where we seek to assess variation in populations for aspects of reaction norms. For

78   example, we may be interested in how much variation in average slope, relative to variation

79   in mean values, occurs among the reaction norms of the genotypes segregating within some

80   population, or among populations within a species. There are two basic approaches in use

81   to quantifying such variation. In the first procedure, two steps are employed. First, data

82   from each genetic group (individual, genotype, inbred line, etc.) are subjected to statistical

83   analysis, for example, to regression analyses to determine slopes, as well as calculations of

84   line-specific means across environments. In the second step of the first approach, summary

85   statistics are calculated at the population level, providing, for example, measures of variance

86   in the means and slopes estimated in the first step. In the second type of approach, mixed

87   models, in particular, random regression mixed models, may be used to directly estimate

88   variance in reaction norm parameters. We show analytically how the two-step approach in-

89   troduces biases into most inferences about variation in reaction norms, and we illustrate the

90   application of random regression mixed models, in detail, with an empirical example. We

91   also derive new measures of variation in phenotype arising from different aspects of reaction

92   norms, and show how these may be particularly useful for answering questions of current

93   interest in reaction norm research.

94   This paper is arrayed in several sections. In each, biologically-relevant results and the

95   more intuitive pieces of statistical theory upon which they rest are presented, while more

96   involved statistical theory is generally relegated to an extensive appendix. First, we consider

97   methods for characterising aspects of individual reaction norms, including the slope between

98   mean phenotype in two environments, and polynomial regression. We then turn our attention

99   to inference of variation in reaction norms. We compare the two-step and mixed model-based

100   approaches, present an empirical example, and derive new measures of variation in reaction

101   norm shape. In the discussion, we recapitulate our major points and address various common

102   threads.

103 **Slopes between two points**

104 The simplest inference of a reaction norm slope is provided by taking the difference between
105 mean phenotype in two environments, for some unit of biological organisation (clone, genetic
106 line, population, species; see for e.g., Berg et al. 2010; Ellers and Driessen 2011; Fallis et al.
107 2014; Liefting et al. 2009). Divided by the difference between the two environments, the
108 difference in mean phenotype gives an estimate of the average slope of the reaction norm
109 between those two points

$$s_{ab} = \frac{\bar{z}_b - \bar{z}_a}{b - a}.$$ (1)

110 This simple assessment of reaction norm slope has two important properties. First, it is an
111 unbiased estimator of the average slope of a reaction norm between points a and b, weighting
112 all values of the environment between a and b equally. The slope of an arbitrary reaction
113 norm function $E[z|x] = f(x)$, where $E[z|x]$ is the expected phenotype, z, given the value of
114 the environmental variable, x, at any given point, and $f'(x)$ is its derivative of the function
115 $f(x)$, at point x. An average over a continuous variable can be obtained by integrating the
116 quantity to be averaged, i.e., $f'(x)$, over the range of the predictor variable (the environment,
117 x between a and b), while weighting by the probability density of x (in this case a uniform
118 density between a and b, which is $\frac{1}{b-a}$), so

$$E[f'(x)] = \int_a^b f'(x)\frac{1}{b-a}dx.$$ (2)

119 Simplifying this expression using basic algebra and calculus rules gives

$$E[f'(x)] = \int_a^b f'(x)\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b f'(x)dx = \frac{1}{b-a}[\int f'(b) - f'(a)] = \frac{\bar{z}_b - \bar{z}_a}{b-a} = s_{ab}.$$ (3)

120 Thus, regardless of the true form of the reaction norm function, i.e., of $f(x)$, the very simple
121 expression in equation 1 gives the average slope of the reaction norm, weighting all values
122 between a an b equally. We will presently see that this ability to recover a major and
123 biologically relevant aspect of a reaction norm is not necessarily a property of other analytical

124 approaches, including some that have recently been advocated in the literature.

125   Second, the basic experimental design associated with the reaction norm analysis in equa-
126 tion 1 can be shown to be optimal with respect to minimising statistical error in the inference
127 of the average slope. If a researcher can rear a set number of individual organisms across
128 a range of environments, it may be desirable for different purposes to raise them in two
129 environments, i.e., at $x = a$ and $x = b$, or to divide the total sample size among additional
130 environments between a and b.

131   It may initially seem that raising organisms across a number of different environments,
132 calculating the slopes between adjacent environmental treatments, and averaging these slopes,
133 would give a better calculation of average slope over some total range of x. This is not the
134 case. The standard error of an estimated reaction norm slope between two points, s_{ab} is

$$SE(s_{ab}) = \sqrt{\frac{\Sigma(\bar{z}_a) + \Sigma(\bar{z}_b)}{b - a}}, \tag{4}$$

135 where $\Sigma(\bar{z})$ denotes the sampling variance of an environment-specific estimate of mean pheno-
136 type, i.e., the squares of the standard errors of the estimated means. The sampling variance
137 of the mean, under normality, is the variance divided by the sample size. The sampling vari-
138 ance of s_{ab} will be minimised when the quantity $\Sigma(\bar{z}_a) + \Sigma(\bar{z}_b)$ is minimised, and if variances
139 are equal in environments a and b, this occurs if the total sample size is divided between the
140 two environments. If variances are not equal in the two environments, a design that increases
141 sample size in the environment with more variance will be optimal for minimising error in
142 s_{ab}.

143   If, alternatively, there were three environments, say x_1, x_2, and x_3, the mean phenotype
144 in environment x_2 would appear in the calculation of s_{ab} for both the intervals between x_1
145 and x_2, and between x_2 and x_3. This produces a negative sampling covariance between the
146 two estimates of s_{ab} for adjacent ranges of x. Consequently, for the purposes of minimising
147 statistical error a single measure of s_{ab} can give the most powerful possible estimate of the
148 average slope of a reaction norm between points $x = a$ and $x = b$. That this design is optimal

149 with respect to minimising the sampling variance of the average slope is demonstrated more

150 rigorously in appendix section A.1.

151 Polynomial regression analysis of reaction norms

152 Polynomial regressions are proven statistical tools for characterising functions, and have

153 been advocated for analysis of non-linear reaction norms (e.g., Gavrilets and Scheiner 1993;

154 Rocha and Klaczko 2014). Polynomial regressions will typically be least-squares fits of an

155 approximating function to a true reaction norm with an unknown true functional form. Given

156 phenotypic values, z, and environmental values, x, for units of observation indexed i, first-

157 (linear), second- (quadratic), and third-order (cubic) polynomial regressions take the form

$$y_i = a + b_1 x_i + e_i,$$

158

$$y_i = a + b_1 x_i + b_2 x_i^2 + e_i,$$

159

$$y_i = a + b_1 x_i + b_2 x_i^2 + b_3 x_i^3 + e_i.$$

160 In each fitted regression model, the intercept, a, and polynomial regression coefficients, i.e., b_1,

161 b_2 and b_3, will be those that minimise the variance of the residuals (e). Note that the values

162 of the intercept and common coefficients (e.g., b_1) may differ between models of different

163 polynomial order, fitted to the same data.

164　In application of polynomial regression, it is hoped that coefficients of the regression

165 model, or predictions from the fitted model, will reflect biologically relevant aspects of re-

166 action norms. While polynomial regression may often be pragmatic, the conditions under

167 which coefficients of polynomial regression models will reflect specific, biologically relevant,

168 quantities such as the average slope of curvature are limited. Where there is a very simple

169 and general interpretation of s_{ab} as the average slope of a true arbitrary reaction norm be-

170 tween the points $x = a$ and $x = b$, the conditions under which a polynomial regression can

171 provide a similar inference are much more limited. If the environmental variable is normally

172 distributed - both in the data analysed and in the relevant scenario in nature about which

173 we want to draw inferences, then the linear term (b_1) in a first-order quadratic regression

174 model gives the average slope. This can be demonstrated from Stein's lemma (Stein, 1973),

175 where it has been shown that $\sigma(xy) = \sigma^2(x)E\left[\frac{\delta y}{\delta x}\right]$ if x is normally distributed, not otherwise.

176 A regression coefficient is the covariance of the predictor and the response, divided by the

177 variance of the predictor, so $b_1 = \frac{\sigma(xz)}{\sigma^2(x)} = E\left[\frac{\delta z}{\delta x}\right]$. The linear term in higher-order polynomial

178 regression models estimates the average slope as well, still requiring normality of the envi-

179 ronmental variable, and also that the environmental variable has a mean of zero. Similarly,

180 if the environmental covariate is normally-distributed and mean-centred, the quadratic term

181 in a quadratic approximation to the reaction norm is equal to half of the average second

182 derivative of the reaction norm function. These specific properties of quadratic regression

183 analysis, when covariates are normal and mean-centred, underlie regression-based analysis of

184 selection gradients, which are the average first and second (partial) derivatives of (relative

185 fitness) functions as well (Geyer and Shaw, 2010; Lande and Arnold, 1983; Mitchell-Olds and

186 Shaw, 1987).

187 So, there is a condition, namely, normality of the environmental covariate x, under which

188 coefficients of polynomial regressions have very general and biologically useful interpreta-

189 tions. Under normality, the slope and curvature of a polynomial approximation to a reaction

190 norm can reflect the average slope and curvature of the true reaction norm, regardless of the

191 true form of the reaction norm. However, the condition of normality of the environmental

192 variable (x) is patently not met in virtually all studies of reaction norms. Rather, by design,

193 the distribution of the environmental variable(s) are non-normal, being composed of two or

194 more discrete treatments. This distribution for x tends toward a uniform distribution as the

195 number of treatments increases. If the distribution of the environmental variable is not nor-

196 mal, then the parameters of a polynomial regression have no direct biological interpretation,

197 and no single useful statistical interpretation, other than that they are the parameters that

198 minimise the residual variance. However, parameters of a polynomial regression will provide

199 insight into aspects of true reaction norms, regardless of the distribution of x, providing they
200 are a reasonably good approximation of the true reaction norm.

201   There is a corollary of the principle that the parameters of a polynomial regression need
202 only reflect aspects of a true reaction norm under specific assumptions about the distribu-
203 tion of the covariate that may appear more biologically important. This is as follows: the
204 parameters of a polynomial approximation to a reaction norm are not just functions of the
205 true reaction norm; but, they are also determined by the distribution of the environmental
206 variable in any given analysis. Figure 1 shows polynomial approximations to an arbitrary
207 (non-polynomial) function, as well as the differences in polynomial approximations to a re-
208 action norm that occur as a result of a near-uniform distribution (i.e., many closely-spaced
209 environmental treatments; figure 1a), and a normal distribution (dashed lines in figure 1b),
210 where both distributions have the same mean and variance. The differences in polynomial
211 shape, arising due to only a change in the distribution of the covariate, are substantial. The
212 quadratic approximation is much more steeply peaked when the distribution of x is uniform,
213 rather than normal. The cubic approximation contains a minimum within the range of the
214 covariate for the uniform distribution, but not for the normal distribution. While the near-
215 uniform covariate distribution (many, closely-spaced environmental treatments) is advocated
216 (Rocha and Klaczko, 2012) and used (e.g., Morin et al. 1999; Pétavy et al. 2001; Rocha et al.
217 2009), in studies of reaction norms, it may often be that extreme environments are relatively
218 rare in nature, and environmental variables may be more normally distributed. Regardless
219 of specific distributions and their relevance in different situations, the fact that polynomial
220 regressions do not reflect only the reaction norm being studied, but also essentially arbitrary
221 features of a study design, should be strong reason for care in their interpretation. In ap-
222 pendix section A.2, we demonstrate a simple example where the s_{ab} metric can be useful,
223 and slopes of quadratic approximations of reaction norms may be less useful.

224   There are likely many scenarios where polynomial regression will provide pragmatic and
225 useful statistical models for studying reaction norms. While the reaction norm used for

226 illustration in figure 1 is very plausible – indeed, this sort of functional form appears in
227 many discussions of thermal reaction norms (e.g., Kingsolver et al. 2004) – many studies
228 will not have to contend with the same degree on non-linearity. As a polynomial regression
229 more closely approximates the true function, predictions from the approximation will better
230 reflect aspects of the biology of the reaction norm. The degree of model complexity, i.e.,
231 the degree of a polynomial reaction norm, is difficult to determine. Previous discussions of
232 polynomial reaction norms have suggested forward model selection. Such a procedure can be
233 inconsistent, i.e., can fail to converge on the "true" model (in the hypothetical situation where
234 the true model is included in the set of models that is considered), even when arbitrarily large
235 amounts of data are available. We elaborate on this property of forward model selection of
236 polynomial regression functions in appendix section A.3.

237 Assessing variation in reaction norms

238 Variation among reaction norms, for example genetic variation among species, families, clones
239 or inbred lines, is often assessed by first calculating metrics such as s_{ab}, or by fitting poly-
240 nomial regression functions, to each genetic unit (e.g., clone, genotype, sibship). In a second
241 step, variances (or other measures of variation) in s_{ab}, or of regression parameters, are cal-
242 culated. This basic procedure will exaggerate apparent levels of variation in any feature of
243 reaction norms, a principle that can be demonstrated with some simple theory about the
244 sampling variance of regression coefficients.

245   The sampling variance of the mean (the intercept in a linear model with a symmetric,
246 mean-centred covariate) is

$$\Sigma[\bar{z}] = \Sigma[\mu] = \frac{\sigma_r^2}{n},$$

247 (the square-root of which is the familiar formula for the standard error of a mean), and the
248 sampling variance of the slopes is

$$\Sigma[b_1] = \frac{3\sigma_r^2}{nr^2}.$$

249 A derivation of this expression is given in the appendix, section A.4.

250   What is the significance of these sampling error variances? In the common two-step pro-
251 cedure, where parameters such as line-specific slopes are first calculated, and then variances
252 (or other summary statistics) of those statistical estimates are subsequently calculated, sta-
253 tistical noise in the first step gets interpreted as biological variation in the second step. The
254 amount of statistical variation that may be interpreted as biological variation in the linear
255 approximation term to a family of reaction norms is thus $\frac{3\sigma_r^2}{nr^2}$. Since the residual variance is
256 always positive, the two-step assessment of variation in reaction norm parameters will always
257 be upwardly biased. Because the number of environments will typically be modest (n is
258 the number of points in the regression, and this is typically the number of environmental
259 treatments), this effect can be large. This effect of statistical error in step 1 to contribute to
260 the apparent variation in step 2 will occur in both inferences of average reaction norm slope
261 in estimates of s_{ab}, and in regression-based approaches.

262   The basic statistical theory that gives sampling errors of regression parameters can give
263 the sampling variance of quadratic terms b_2 in the notation introduced above. The sampling
264 error of quadratic terms given a centred uniform covariate is

$$\Sigma[b_2] = \frac{45\sigma_r^2}{4nr^4}.$$

265 See the appendix section A.4 for a more detailed derivation. Comparison of the expressions
266 for sampling variances of the mean, linear, and quadratic terms, i.e., of $\Sigma[\bar{z}]$, $\Sigma[b_1]$ and $\Sigma[b_2]$,
267 reveals a further complication arising in the two-step procedure for inferring variation in
268 reaction norm parameters. The relative contribution of statistical noise to apparent variation
269 in means and linear and quadratic terms (and higher terms pertaining to other aspects of
270 reaction norm curvature) varies depending on the arbitrary scaling of the covariate: the three
271 expressions for sampling variance are different functions of the parameter r, the essentially
272 arbitrary range of the environmental variable.

273   The two important points, (1) that statistical noise will be interpreted as biological vari-

274   ation in two-step analytical procedures, and (2) that the extent to which statistical noise

275   pollutes biological inferences depends on scaling, are not artefacts of the simplifying assump-

276   tions made here. For example, it occurs if the covariate is not strictly uniform, but rather is

277   composed of few or many distinct environmental treatments. The pattern will also hold for

278   very different distributions of the covariate; analogous expressions for $\Sigma[\mu]$, $\Sigma[b_1]$, and $\Sigma[b_2]$,

279   given a normal covariate, x, are given in appendix section A.5.

280   A class of linear mixed models called random regression models exists specifically to sep-

281   arate noise from real variation in families of regression coefficients. The simplest random

282   regression mixed model is a linear random slopes model, which can be written as

$$z_{ij} = a + bx_i + f_j + g_j x_i + e_i, \tag{5}$$

283   where z_{ij} is the phenotypic observation of individual i from group (e.g., species, family, etc.)

284   j, where a and b are fixed regression parameters for the intercept and slopes, respectively,

285   where f_j and g_j are regression parameters (contrasts to a and b) for group j, x_i is the

286   environment to which individual i was exposed and e_i is a residual for individual i. As

287   before, the residuals are assumed to be drawn from a normal distribution $e_i \sim N(0, \sigma_r^2)$, and

288   furthermore, the group-specific regression parameters are also treated as random variables,

289   i.e., variables that belong to a bivariate normal distribution

$$\begin{bmatrix} f \\ g \end{bmatrix}_j \sim N \left(\mathbf{0}, \begin{bmatrix} \sigma^2(f) & \sigma(f,g) \\ \sigma(f,g) & \sigma^2(g) \end{bmatrix} \right)$$

290   such that $\begin{bmatrix} \sigma^2(f) & \sigma(f,g) \\ \sigma(f,g) & \sigma^2(g) \end{bmatrix}$ is a matrix containing the variances and covariances of slopes

291   and intercepts. Solutions to the mixed model give estimates of terms including the variance

292   in slopes ($\sigma^2(g)$), that are not inflated by sampling error, as occurs in the two-step approach.

293   While random regression analysis is currently in use (Dingemanse et al., 2010; Martin et al.,

294   2011), we hope that it is useful to clarify that its use represents more than a mere modernisa-

295 tion of statistical approaches to studying reaction norms. Random regression can yield direct

296 inferences of variation in reaction norm parameters (e.g., of slopes), that would otherwise be

297 subjected to potentially biologically misleading statistical biases in two-step analyses.

298   A random regression model can be used to assess variation in reaction norms for any

299 analysis with multiple units of observation, and two or more (or a continuous range of)

300 environmental treatments. When applied to a study with two treatments, the linear random

301 regression mixed model specified by equation 5 yields unbiased estimates of the among-group

302 variance in intercepts, average slopes (i.e., this amounts to an analysis of variation in s_{ab}),

303 and their covariance.

304   When applied to a study with a range of environmental conditions, or with random

305 quadratic (or even higher order) terms, random regression mixed model analysis can be

306 used to recover meaningful information about variation in reaction norm shape. However the

307 caveats that apply to the interpretation of polynomial approximations to reaction norms in

308 general will also apply to inferences about variation in polynomial coefficients obtained by

309 random regression. With prudence, it is possible that random polynomial regression mixed

310 model analysis could be much more extensively used in analysis of variation in reaction

311 norms, and such analysis will certainly be preferable to two-step analytical approaches in

312 most circumstances.

313 **Example application of a random regression mixed model**

314 We applied quadratic random regression mixed model analysis to the data on reaction norms

315 reported in Rocha et al. (2009) and re-analysed in Rocha and Klaczko (2012) (data provided

316 by F. B. Rocha and L. B. Klaczko). The data consist of 1122 *Drosophila mediopunctata*

317 phenotyped for abdominal spot number and thorax length, raised in three simultaneous

318 replicates (vials) in a thermal gradient spanning 14°C to 24°C in 1°C intervals. For each

319 trait, the mixed model took the form

$$z_{i,j} = a + b_1 t_i + b_2 t_i^2 + sex_i + f_j + g_{1,j} t_i + g_{2,j} t_i^2 + replicate_i + T_i + e_i, \tag{6}$$

320 where $z_{i,j}$ represents the phenotype (spots or thorax length) measured on individual i belong-

321 ing to strain j. The fixed effects, a, b_1, and b_2 estimate the average reaction norm, conditional

322 on a fixed effect of sex. The random polynomial coefficients f_j, $g_{1,j}$, and $g_{2,j}$ for each line

323 and are assumed to be drawn from a multivariate normal distribution

$$\begin{bmatrix} f \\ g_1 \\ g_2 \end{bmatrix}_j \sim N\left(\mathbf{0}, \mathbf{\Sigma}\right), \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma^2(f) & \sigma(f, g_1) & \sigma(f, g_2) \\ \sigma(f, g_1) & \sigma^2(g_1) & \sigma(g_1, g_2) \\ \sigma(f, g_2) & \sigma(g_1, g_2) & \sigma^2(g_2) \end{bmatrix},$$

324 with estimated covariance matrix $\mathbf{\Sigma}$. Additionally, the replicate associated with individual i,

325 and the temperature in which it was raised, coded as a multi-level factor t_i and the residuals,

326 e_i, are all included as random effects with estimated variances. The temperature at which a

327 given individual was raised, t_i was mean-centred by subtracting 19°C.

328   The among-line covariance matrices of intercepts, slopes, and curvatures (table 1) are

329 difficult to interpret directly. However, some features of the mixed model analysis are imme-

330 diately apparent. First, we can see that, as predicted by the statistical theory given above,

331 the variance of coefficients of reaction norms (table 1d) in the two-step procedure inflates

332 the apparent amount of variation in reaction norm parameters. Another such comparison

333 yielding similar inflation of apparent variation in reaction norm parameters is reported in

334 Liefting et al. (2015). This effect is larger in cases when reaction norms are more similar, in

335 this case, with a much more dramatic effect for thorax length reaction norms than for spot

336 number. Furthermore, the correlations among reaction norm parameters are consistently

337 smaller in inferences from the two-step procedure. This is because statistical noise inflates

338 all of the estimates of variance in the polynomial coefficients, but not necessarily all of the

339 covariances.

340   However, biological inference based on the estimated variances and covariances of polyno-
341   mial reaction norm coefficients is difficult. This is because the relationship between reaction
342   norm shape and slopes and intercepts depends on the scaling of the environmental covariate
343   (temperature, in this case). A first step to interpreting the mixed model results might be to
344   visualise the family of reaction norms implied by the fitted mixed model. Figure 2 shows the
345   raw means for each line in each environment (a and b), quadratic regressions fitted for each
346   line (c and d), and an example of 20 reaction norms simulated from the values of the fitted
347   mixed models (e and f). The last depictions are essentially simulations from the inferred
348   distribution of reaction norms, generated by drawing intercepts, linear and quadratic terms
349   from a multivariate normal distribution with a mean defined by the fixed effects in the fitted
350   model, and with (co)variances set to those estimated by the random effects (table 2).

351   While these reaction norms were previously interpreted as showing ubiquitous effects of
352   variation in reaction norm shape, this interpretation seems tenuous based on consideration of
353   the visualisations of the families of reaction norms in figure 2. For spot number, all inferences,
354   including those that inflate the amount of variation in reaction norm shape (parts a and c),
355   indicate that the reaction norms are approximately linear and thus there is in fact only very
356   modest variation in reaction norm shape. A mixed model analysis is particularly useful
357   for separating shared features of reaction norms (characterised by fixed effects) from ways
358   that they vary (characterised by random effects). Indeed, the previous interpretation that
359   these reaction norms show that curvature is a common feature of reaction norms (Rocha
360   and Klaczko, 2012) is true. However, for thorax length, a critical further finding is that
361   the reaction norms of different lines have very similar curvatures; this allows more nuanced
362   interpretation of when and how curvature is an important feature of reaction norms.

363   For thorax length, variation in the reaction norms is even more modest. The crossing of
364   reaction norms has been suggested as a measure of variation in slope, and multiple crossing of
365   reaction norms as a measure of variation in curvature (Rocha and Klaczko, 2012). However, if
366   reaction norms vary very little in any way, then a great deal of crossing occurs! Consider figure

367 2b; if all of the reaction norms were nearly identical, then statistical noise in estimating the

368 mean phenotype for each strain in each environment would cause half of the line segments in

369 a plot such as this to cross. Clearly, line-crossing is difficult to apply as a measure of reaction

370 norm complexity.

371 Variance in phenotype arising from variation in reaction norm parameters

372 How can we make inferences about the relative importances of variation in the mean values

373 of reaction norms, and of slopes and quadratic terms, if the variances of intercepts, slopes,

374 and quadratic terms depend on the arbitrary scaling and distribution of the environmental

375 covariate (and additionally on the covariance of intercepts and quadratic terms)? Given any

376 distribution of the environmental covariate, it is possible to derive the amount of variation

377 among genetic units (e.g., lines), integrated over the distribution of the covariate, that arises

378 from differences in the mean, slope and quadratic curvature of reaction norms. Detailed

379 derivations are given in the appendix (section A.6 for a uniform covariate and A.7 for a

380 normal covariate), and the formulae for these measures of variation in different aspects of

381 reaction norm shape are given in table 2.

382   Figure 3 shows the amount of variation in expected line- and temperature-specific pheno-

383 typic values for both traits in the *Drosophila* example that are attributable to variation in the

384 means, slopes, and curvatures of the families of reaction norms, as assessed by the quadratic

385 random regression mixed model analysis (equation 6), and by the two-step analytical pro-

386 cedure. This quantification of different components of variation in reaction norms confirms

387 that differences in reaction norm shape are indeed modest, and also further demonstrates

388 the danger of inflating inferences in the two-step analytical procedure (figure 3a,b). The

389 majority of variation among lines arises from differences in mean values of reaction norms.

390 Calculations of variation attributable to intercepts, linear, and quadratic terms, separately

391 made based on uniform and normal distributions, are quite similar in this example (compare

392 figures 3a,b with 3c,d). This should generally be the case when families of quadratic functions

393   capture variation in reaction norms well, as seems to be the case for these data.

394   (Co)variances of arbitrary reaction norm properties

395   Intercepts, linear, and quadratic terms, and the variance in expected values with which they

396   are associated, do not directly represent all features of reaction norms in which we may be

397   interested. For example, we may be interested in phenotypic values at specific environmen-

398   tal values, mean phenotypic values integrated over different distributions of environmental

399   values, locations of maxima or minima (environments that produce minimum or maximum

400   phenotypes), and phenotypic values at maxima or minima (minimum or maximum phenotypic

401   values). Quadratic regressions contain information about such reaction norm properties, pre-

402   viously sometimes termed "characteristic values" (Delpuech et al., 1995; Gibert et al., 1998),

403   and families of regression coefficients, as estimable by random regression analysis, contain

404   information about means and variances of such reaction norm properties. Operationally,

405   calculations of variance in arbitrary reaction norm properties seems easiest in the two-step

406   analytical procedure. In mixed model analyses, one must call on somewhat more statisti-

407   cal sophistication to derive (co)variances of reaction norm properties from estimated means,

408   variances, and covariances of quadratic reaction norm parameters. However, an approach

409   to develop formulae for such quantities seems clear. Given a function for calculating some

410   quantity (e.g., the location of an optimum) from a fitted reaction norm function, the vari-

411   ance of that quantity can be approximated by taking the expectation of a Taylor series. It

412   is reasonable to feel that this is easier said than done. It may therefore be useful to provide

413   expressions involving some quantities that might be most useful.

414   Table 3 gives expressions for quantities that may be calculated from quadratic reaction

415   norm approximations: the environment of the maximum or minimum value, the maximum or

416   minimum value, and the mean (remembering that the intercept is not the mean of a quadratic

417   function, even if the covariate is mean centred) for different distributions of the environmental

418   covariate. Table 3 also gives expressions for the expectations and variances of each of these

419 quantities, given means and variances and covariances of quadratic regression coefficients, as
420 are obtained from random regression mixed model analysis. Similar approaches to those that
421 yielded these expressions (given in appendix section A.8), can be used to give variances and
422 covariances of multiple derived reaction norm properties.

423   We can briefly explore the application methods to infer distributions of arbitrary reaction
424 norm properties using the data on the reaction norms of thorax length as a function of
425 temperature (table 1, figure 2). While mean values for spot number may be biologically
426 informative, means and variances of locations and values of optima for spot number will
427 not. This is because the distribution of reaction norms contains very many nearly linear
428 functions (figure 2), and optima of the quadratic approximations of such functions are far
429 from the relevant range of temperature. However, the distribution of reaction norms for
430 thorax length appears to have a reasonably well-defined maximum, and knowing how this
431 maximum's value and location varies among reaction norms may be of biological interest.
432 The mean and variance of locations of maxima for thorax length, as given by the expressions
433 in table 3 are 15.7 and 1.11, respectively. Means and variances of the phenotypic values at
434 the maxima are 1.56 and 2.93^{-4}, respectively.

435   The approach to obtaining expressions for (approximating) the distributions of arbitrary
436 reaction norm properties, as given in table 3 and appendix section A.8, could be extended in
437 order to obtain other metrics of potential interest, for example, the covariance of locations
438 and values of optima; however, the expressions will become increasingly unwieldy. The main
439 value of the expressions given in table 3 is that they demonstrate that the random regres-
440 sion approach provides the information necessary to infer arbitrary properties of families of
441 reaction norms, given the assumption that a family of quadratic functions gives a reasonable
442 approximating model. A more pragmatic option is available. Monte-Carlo (MC) simulation
443 can provide very precise approximations to quantities such as means and variances of arbi-
444 trary quantities. In the present setting, the technique would require simulation of a large
445 number (say a million) random normal vectors with means and covariances equal to those

446 estimated by the fixed and random parts of a quadratic random regression mixed model.
447 Then for each simulated vector, one or more quantities of interest can be calculated. The
448 means, variances, and covariances of these simulated reaction norm properties will approach
449 the values defined by the fitted random regression mixed model, as the number of simulations
450 becomes large. On modern personal computers, this kind of procedure takes seconds. By
451 this procedure (with one million MC simulations) the mean of the environments of maximum
452 values, and the mean of the maxima, are 15.6 and 1.56. The variance of environments of
453 maximum values is 1.29, the variance of maximum values is 3.18^{-4}, and the covariance of the
454 two quantities is -1.01^{-2} (the associated correlation is -0.497). The first four of these statistics
455 agree very closely with the four values given above based on approximations given in table 3,
456 indicating that the approximations may generally be robust. Further, it may be of particular
457 interest in a biological example such as this, that those lines with the highest optima have
458 their optima at the lowest temperatures. MC simulation approaches can be applied as well
459 to obtaining standard errors of statistics of the distributions of arbitrary quantities.

460   Another property of sets of reaction norms that may be of biological interest is the
461 environment-specific variance of traits, and covariance among genetic units across environ-
462 ments. Any parameters describing the variances and covariance of reaction norm parameters
463 (e.g., intercepts, slopes, etc.) defines a specific pattern of within- and across-environmental
464 (co)variance in phenotype. We have focused so far on a reaction norm, or "function-valued
465 trait" approach to phenotypic plasticity, but this perspective is entirely complimentary to
466 thinking about environment-specific covariances, which is sometimes called the "character
467 state" approach (van Tienderen and Koelewijn, 1994). For example, in Box 1 figure B1a,
468 a we could make a character state description of the reaction norms by saying that there is
469 little variance in the trait associated with genotypes at $x = 4$, but appreciable variance for
470 higher an lower values of the environmental variable; furthermore, we could state that, at
471 the genetic level, trait values when $x < 4$ are negatively correlated with trait values when
472 $x > 4$. A character state representation can sometimes be a very useful way of describing the
473 properties of a family of covarying reaction norm parameters. Box 2 provides a description

474 of how to represent covariances of reaction norm parameters as environment-specific means
475 and covariances.

476   The quantities discussed in this and the previous section for summarising reaction norms
477 apply directly to families of quadratic regressions (and associated formulae in tables 2 and
478 3), especially as can be estimated with random regression approaches in linear mixed models.
479 The reasoning behind these could in principle be extended to other types of functions, for
480 example, to higher-order polynomials. In analyses of non-normal traits, e.g., with random
481 regression in generalised mixed models, quantities described here would apply on the under-
482 lying latent scale (e.g., on the log scale in a Poisson model), which in many cases could be
483 very useful and biologically interpretable.

484 Discussion and Conclusions

485 We are neither advocating for, nor against, characterising reaction norm slope as the slope
486 of a line between two points, i.e., of s_{ab}, on the reaction norm. We seek primarily to clarify
487 that this very simple statistic has a very specific interpretation (i.e., the average slope of the
488 reaction norm between points a and b) that holds regardless of the true shape of the reaction
489 norm. Few other statistics one might use to characterise reaction norms have interpretations
490 that holds so generally. This does not mean that s_{ab} could not potentially be misleading. For
491 example, if a reaction norm contains a minimum or a maximum between the points $x = a$
492 and $x = b$, then it is difficult to see what biological use inferences of s_{ab} may be, without fur-
493 ther detailed analyses of reaction norm shape. Thus, there are situations where s_{ab} contains
494 exactly the information that is needed, even if a reaction norm is highly non-linear, and there
495 are times when information is needed that s_{ab} cannot provide. Understood correctly, there is
496 neither a "pessimistic" nor an "optimistic" view (Rocha and Klaczko, 2012) to be had about
497 s_{ab}; rather it is fit for some purposes and irrelevant to others. Additionally, it is of note that
498 the distribution of slopes and intercepts of straight-line reaction norms between two environ-
499 ments has direct relationships to cross-environment phenotypic and genetic correlations (Via

500 and Lande, 1985), another simple and robust relationship that does not necessarily hold for
501 more complex statistical models of reaction norms.

502   Similarly, while we have attempted to be very careful about the narrowness of the con-
503 ditions under which coefficients of polynomial regressions can be interpreted as reflecting
504 specific properties of true reaction norms, we are neither advocating for, nor against, poly-
505 nomial regression. In particular, we note that polynomial approximating functions depend
506 on the distribution of the covariate (i.e., the environmental variable), and cannot therefore
507 necessarily be interpreted solely as properties of reaction norms. More importantly, we have
508 noted that the slope of an approximation to a reaction norm at any specific point (i.e., Rocha
509 and Klaczko 2012's "local plasticity") is not necessarily a good representation of the slope
510 of the true reaction norm at that point. If a polynomial regression is sufficiently flexible
511 relative to the presumed complexity of a true reaction norm, then this technique is likely to
512 be valuable, even if it does not generally yield estimates of quantities that have such a simple
513 and general interpretation as estimates of s_{ab}. We suspect that quadratic regression could
514 prove to be a very pragmatic model of many non-linear reaction norms.

515   Despite the previous claim that the *Drosophila* data in figure 2 support a contention that
516 reaction norms vary extensively in their curvature, it is fairly easy to see that this is not
517 the case. Taking into consideration that finite sample size for each line in each assayed
518 temperature causes statistical noise in each point in figure 2a,b, it is clear that there is very
519 little variation in reaction norm shape among lines. For abdominal spot number, the raw
520 data consist primarily of parallel lines. For thorax length, dispassionate visual inspection and
521 quantitative analysis shows that most of the variation is explained by the overall reaction
522 norm, and by variation within temperatures, not differences in reaction norms among lines.
523 Note that quantitative approaches (table 1, figure 5) support these contentions based on
524 the raw data. Since two-step analytical procedures have been widely used in the primary
525 literature, and even meta-analysis (Murren et al., 2014), it is not currently possible to judge
526 how flexible polynomial regressions might generally have to be to capture the most important

527 features of (variation in) reaction norms. However, the analyses here are heartening and it

528 seems plausible, that with due consideration to the features of any particular study system,

529 that quadratic regressions, as advocated by (Rocha and Klaczko, 2014), could indeed provide

530 pragmatic models of reaction norms in many cases.

531 However, we caution strongly against some of Rocha and Klazcko's (2014) specific sugges-

532 tions for interpreting quadratic regressions. In particular, Rocha and Klazcko suggest that

533 the derivative of a quadratic, or other polynomial function at any specific point, which they

534 call "local reaction norm plasticity" could be a generally useful measure of reaction norm

535 shape at a particular point. However, this derivative need not necessarily closely reflect the

536 slope of a true reaction norm at that point, and it need not even be the correct sign (see also

537 appendix section 2, figures A.1 and A.2). Rocha and Klaczko (2014) also suggest that the

538 quadratic term can be used as a measure of "reaction norm shape", justified by the fact that

539 twice the value of the quadratic term is the second derivative of the quadratic function at

540 all points, and is therefore the average derivative of the quadratic function. This use seems

541 reasonable, but its application should be approached with awareness that the average deriva-

542 tive or second derivative of the quadratic approximation to any function is only equal to the

543 average derivative or second derivative under two conditions. First, this equality holds if the

544 true function is indeed quadratic. Second, this equality holds if the environmental covariate

545 is normal. Virtually no studies of reaction norms have a normal covariate. In fact, investi-

546 gators typically strive for covariate distributions that approach uniformity. So justification

547 for using curvature of a quadratic approximation as a measure of the curvature of a reaction

548 norm rests on a requirement that a quadratic function is a good approximation of the true

549 reaction norm. It seems that this requirement should frequently be closely enough met in

550 empirical systems for quadratic regression to provide useful measures of average reaction

551 norm curvature.

552 We do advocate more strongly for analyses that do not apply statistical procedures to

553 the outcomes of previous statistical procedures. Doing statistics on statistics will often lead

554 to misleading results, and is generally avoidable. Although some authors have begun to

555 use mixed model analysis (e.g., Dingemanse et al. 2010; Martin et al. 2011), multi-step ap-

556 proaches are still common empirical practice. Furthermore, while the applicability of random

557 regression has been clearly demonstrated (Martin et al., 2011), we hope it is useful to clarify

558 that it doesn't merely represent a modernisation of statistical approaches to studying reaction

559 norms. Rather, it allows pitfalls of two-step procedures to be avoided. When population-level

560 variation is inferred from the distribution of summary statistics of units of organisation that

561 have been analysed within that population, the apparent variation at the population-level

562 will invariably be upwardly biased by statistical noise (see also Morrissey 2016). Given the

563 existence of random regression mixed models in widely-used software packages (e.g., LME4,

564 Bates et al. 2014; MCMCGLMM, Hadfield 2010; ASREML, Gilmour et al. 2002) their use in

565 studies of reaction norms should probably be extended. However, just as the biological utility

566 of polynomial approximations to reaction norm functions depends on the closeness to which

567 they approximate true reaction norm functions, inferences from random regression mixed

568 models will also depend on the adequacy of a family of polynomial functions to describe

569 variation in reaction norms.

570   Additional benefits of mixed model-based analyses that we realised in our example anal-

571 ysis also contribute to the utility of these methods. For example, we were able to account

572 for covariates (by fitting a fixed effect for sex), and possible sources of variation and non-

573 independence among observations (by treating variation among replicate vials as random

574 effects). Furthermore, mixed-model analyses can very naturally account for unequal num-

575 bers of observations in different environmental treatments, while such accounting is more

576 difficult in the two-step procedure (weighting by precision would be necessary, and it is not

577 clear if such an effort has ever been made in reaction norm studies). These seem like valuable

578 aspects of the analysis for ensuring the most complete and rigorous use of the available data,

579 and are relatively easily implemented in the mixed model framework. Furthermore, when

580 a low-order polynomial function does not fit a reaction norm well, mixed model analysis

581 may provide simple and powerful solutions. For example, the average reaction norm may be

582 handled with the fixed effects part of a model, using a high-order polynomial or some other
583 flexible regression function such as a spline regression (Wood, 2006), while a relatively low-
584 order model, such as a quadratic random regression, may still be pragmatic for describing
585 variation in reaction norms around the average function.

586 Linear mixed models, in particular the random regression mixed models considered here,
587 are among the simplest of types of hierarchical model that may be useful to analysis of
588 reaction norm shapes. Non-linear mixed models, and hierarchical models in general, could
589 potentially be used to provide direct inference of variation in parameters such as the locations
590 of maxima, and for coefficients of reaction norm models that are not based on polynomials.
591 Flexible models that can provide such inferences are becoming increasingly easy to implement,
592 for example with software such as JAGS (Plummer, 2010) and STAN (Stan Development Team,
593 2014). With this range of options for model-based inference of variation in reaction norms, it
594 should be increasingly possible to design powerful studies of interesting aspects of phenotypic
595 plasticity.

Acknowledgements

References

603 Bates, D., M. Maechler, B. Bolker, and S. Walker, 2014. lme4: Linear mixed-effects models
604   using Eigen and S4, R package version 1.1-7 ed.

605 Berg, M. P., E. T. Kiers, G. Driessen, M. van der Heijden, B. W. Kooi, F. Kuenen, M. Liefting,

606   H. A. Verhoef, and J. Ellers. 2010. Adapt or disperse: understanding species persistence

607   in a changing world. Global Change Biology 16:587–598.

608 Chevin, L. M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a

609   changing environment: towards a predictive theory. PLOS Biology 8:e1000357.

610 Delpuech, J.-. M., B. Moreteau, J. Chiche, E. Pla, J. Vouidibio, and J. R. David. 1995. Phe-

611   notypic plasticity and reaction norms in temperate and tropical populations of *Drosophila*

612   *melanogaster*: ovarian size and developmental temperature. Evolution 49:670–675.

613 DeWitt, T. J. and S. M. Scheiner, 2004. Phenotypic plasticity - Functional and Conceptual

614   Approaches. Oxford University Press, Oxford.

615 Dingemanse, N. J., A. J. Kazeem, K. Réale, and J. Wright. 2010. Behavioural reaction

616   norms: animal personality meets individual plasticity. Trends in Ecology and Evolution

617   25:81–89.

618 Ellers, J. and G. Driessen. 2011. Genetic correlation between temperature-induced plasticity

619   of life-history traits in a soil arthropod. Evolutionary Ecology 25:473–484.

620 Fallis, L., J. Fanara, and T. Morgan. 2014. Developmental thermal plasticity among

621   *Drosophila melanogaster* populations. Journal of Evolutionary Biology 27:557–564.

622 Gavrilets, S. and S. M. Scheiner. 1993. The genetics of phenotypic plasticity. V. Evolution

623   of reaction norm shape. Journal of Evolutionary Biology 6:31–48.

624 Geyer, C. J. and R. G. Shaw, 2010. Aster models and the Lande-Arnold beta. Tech. rep.,

625   University of Minnesota.

626 Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus

627   non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new

628   environments. Functional Ecology 21:394–407.

629 Gibert, P., B. Moreteau, J. R. David, and S. M. Scheiner. 1998. Describing the evolution of
630   reaction norm shape: body pigmentation in *Drosophila*. Evolution 52:1501–1506.

631 Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson, 2002. ASReml
632   user guide release 1.0. VSN International Ltd, Hemel Hempstead, United Kingdom.

633 Gupta, A. P. and R. C. Lewontin. 1982. A study of reaction norms in natural populations
634   of *Drosophila pseudoobscura*. Evolution 36:934–948.

635 Hadfield, J. 2010. MCMC methods for multi-response generalized linear mixed models: The
636   MCMCglmm R package. Journal of Statistical Software 33:1–22.

637 Kingsolver, J. G., R. Izem, and G. J. Ragland. 2004. Plasticity of size and growth in fluctuat-
638   ing thermal environments: comparing reaction norms and performance curves. Integrative
639   and Comparative Biology 44:450–460.

640 Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection,
641   and evolution of growth trajectories. Genetics 124:979–993.

642 Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters.
643   Evolution 37:1210–1226.

644 Liefting, M., R. H. A. van Grunsven, M. B. Morrissey, M. J. T. N. Timmermans, and
645   J. Ellers. 2015. Interplay of robustness and plasticity of life-history traits drives ecotypic
646   differentiation in thermally distinct habitats. Journal of Evolutionary Biology 28:1057–
647   1066.

648 Liefting, M., A. A. Hoffmann, and J. Ellers. 2009. Plasticity versus environmental canaliza-
649   tion: population differences in thermal responses along a latitudinal gradient in *Drosophila*
650   *serrata*. Evolution 63:1954–1963.

651 Martin, J. G. A., D. H. Nussey, A. J. Wilson, and D. Réale. 2011. Measuring individual
652   differences in reaction norms in field and experimental studies: a power analysis of random
653   regression models. Methods in Ecology and Evolution 2:362–374.

654 Mitchell-Olds, T. and R. G. Shaw. 1987. Regression analysis of natural selection: statistical
655   inference and biological interpretation. Evolution 41:1149–1161.

656 Morin, J. P., B. Moreteau, G. Pétavy, and J. R. David. 1999. Divergence of reaction norms
657   of size characters between tropical and temperate populations of *Drosophila melanogaster*
658   and *D. simulans*. Journal of Evolutionary Biology 12:329–339.

659 Morrissey, M. B. 2016. Meta-analysis of magnitudes, differences, and variation in evolutionary
660   parameters. Journal of Evolutionary Biology in press.

661 Murren, C. J., H. J. Maclean, S. E. Diamond, U. K. Steiner, M. A. Heskel, C. A. Handels-
662   man, C. K. Ghalambor, J. R. Auld, H. S. Callahan, D. W. Pfennig, R. A. Relyea, C. D.
663   Schlichting, and J. G. Kingsolver. 2014. Evolutionary change in continuous reaction norms.
664   The American Naturalist 183:453–467.

665 Nussey, D. H., E. Postma, P. Gienapp, and M. E. Visser. 2005. Selection on heritable
666   phenotypic plasticity in a wild bird population. Science 310:304–306.

667 Pétavy, G., J. R. David, P. Gilbert, and B. Moreteau. 2001. Viability and rate of development
668   at different temperatures in *Drosophila*: A comparison of constant and alternating thermal
669   regimes. Journal of Thermal Biology 26:29–39.

670 Plummer, M., 2010. JAGS version 2.0 Manual. International Agency for Research on Cancer.

671 Rocha, F. B. and L. B. Klaczko. 2012. Connecting the dots of nonlinear reaction norms
672   unravels the threads of genotype-environment interaction in *Drosophila*. Evolution 66:3404–
673   3416.

674 ———. 2014. Undesirable consequences of neglecting nonlinearity: response to comments
675   by Liefting et al. (2013) on Rocha and Klaczko (2012). Evolution 68:1548–1551.

676 Rocha, R. B., H. F. Medeiros, and L. B. Klaczko. 2009. The reaction norm for abdominal
677   pigmentation and its curve in *Drosophila mediopunctata* depend on the mean phenotypic
678   value. Evolution 63:280–287.

679 Scheiner, S. M. 1993. Genetics and evolution of phenotypic plasticity. Annual Review of
680   Ecology, Evolution and Systematics 24:35–68.

681 Scheiner, S. M. and H. S. Callahan. 1999. Measuring natural selection on phenotypic plas-
682   ticity. Evolution 53:1704–1713.

683 Schmalhausen, I. I., 1949. Factors of Evolution. Blakinston, Philedelphia, PA.

684 Stan Development Team, 2014. Stan: A C++ Library for Probability and Sampling, Version
685   2.5.0.

686 Stein, C. M. 1973. Estimation of the mean of a multivariate normal distribution. Proceedings
687   of the Prague Symposium on Asymptotic Statistics 1:345–381.

688 van Tienderen, P. H. and H. P. Koelewijn. 1994. Selection on reaction norms, genetic
689   correlations, and constraints. Genetical Research 64:115–125.

690 Via, S. and R. Lande. 1985. Genotype-environment interaction and the evolution of pheno-
691   typic plasticity. Evolution 39:505–522.

692 West-Eberhard, M. J., 2003. Developmental plasticity and evolution. Oxford University
693   Press, Oxford.

694 Woltereck, R. 1909. Weitere experimentelle Untersuchungen über Artveränderung, speziel
695   uber das Wesen quantitativer Artunterschiede bei Daphniden. Verhandlungen der
696   Deutschen Zoologischen Gesellschaft 19:110–173.

697 Wood, S. N., 2006. Generalized Additive Models: An Introduction with R. Chapman and
698   Hall/CRC.

Table 1: Random regression mixed model-based inference of variation in reaction norms for spot number and thorax length in *Drosophila mediopunctata*, (a) fixed effect estimates, (b) standard deviations of random intercepts, linear, and quadratic terms, and their correlations, and (c) standard deviations of additional random effects. For comparison, standard deviations and correlations of intercepts, linear, and quadratic terms from a traditional multi-step procedure are given in part (d).

	number of spots			thorax length		
(a) fixed effects						
a	1.8			1.54		
b_1	-0.162			-1.21^{-2}		
b_2	-4^{-4}			-1.9^{-3}		
sex	0.243			-0.156		
(b) random quadratic regression coefficients (as SDs and correlations)						
	f	g_1	g_2	f	g_1	g_2
f	0.618			1.54^{-2}		
g_1	-0.263	0.0485		0.117	2.3^{-3}	
g_2	-0.979	0.454	1.35^{-2}	-0.243	-0.992	2^{-4}
(c) additional random variance components (as SDs)						
replicate	0.158			0.0212		
temperature	0.0382			6.6^{-3}		
residual	0.477			0.0571		
(d) SDs and correlations of polynomial coefficients from the two-step procedure						
	f	g_1	g_2	f	g_1	g_2
f	0.642			0.0299		
g_1	-0.0802	0.0447		-0.159	7^{-3}	
g_2	-0.933	0.286	1.69^{-2}	-0.671	-1.25^{-2}	1.2^{-3}

Table 2: Expressions for variance in environment-specific expected values of phenotype attributable separately to variation in reaction norm means, slopes, and curvatures (quadratic terms), for uniform and normal environmental covariates. All expressions assume that covariates are mean-centred. For the uniform distribution, r represents the range, i.e., where the centred uniform covariate has a range from $-r$ to $+r$. For the normal covariate, the environmental variable's distribution is characterised by the standard deviation, σ.

	uniform	normal
mean (Σ_μ)	$\frac{1}{9}\left(6r^2\sigma(f,g_2) + 9\sigma^2(f) + r^4\sigma^2(g_2)\right)$	$\sigma^2(f) + 2\sigma(f,g_2)\sigma^2(x) + \sigma^2(g_2)\sigma^4(x)$
slopes (Σ_{g_1})	$\frac{1}{3}r^2\sigma^2(g_1)$	$\sigma^2(g_1)\sigma^2(x)$
curvatures (Σ_{g_2})	$\frac{4}{45}r^4\sigma^2(g_2)$	$2\sigma^4(x)\sigma^2(g_2)$

699

700

701

702

703

Table 3: Expressions for some properties of quadratic reaction norms, and means and variances of these quantities for families of quadratic reaction norms.

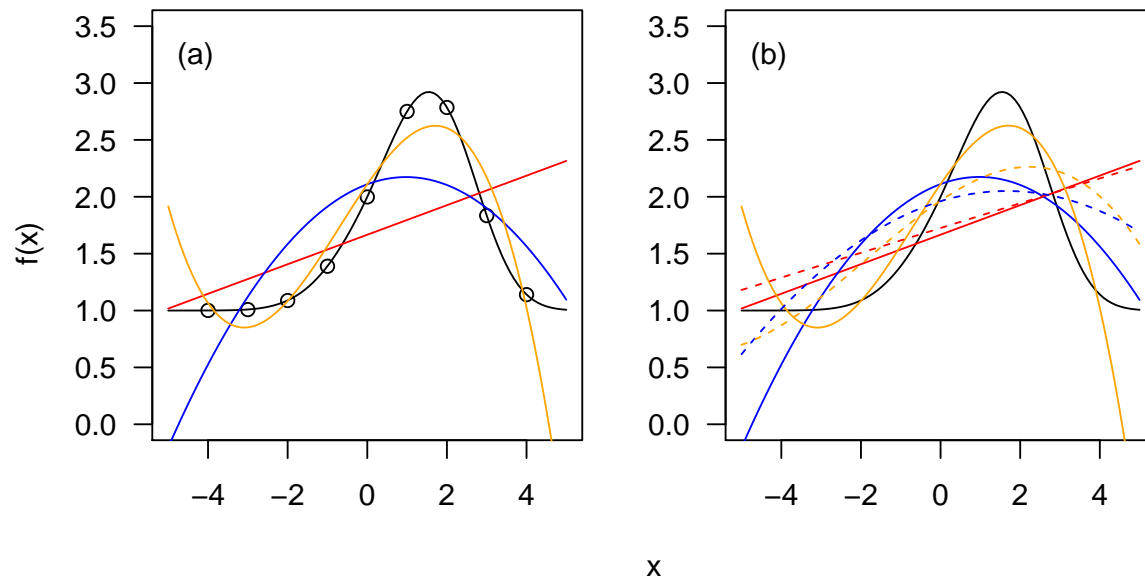| quantity	distribution	quantity	mean value	variance
mean across environments	$N\left(0, \sigma^2(x)\right)$	$a + b_2\sigma^2(x)$	$a + b_2\sigma^2$	$\sigma^2(f) + \sigma^2(g_2)\sigma^4(x) + 2\sigma(f,g_2)\sigma^2(x)$
mean across environments	$U\left(-r, +r\right)$	$a + \frac{b_2 r^2}{3}$	$a + \frac{b_2 r^2}{3}$	$\sigma^2(f) + \sigma^2(g_2)\frac{r^4}{9} + 2\sigma(f,g_2)\frac{r^2}{3}$
average slope	any	b_1	b_1	$\sigma^2(b_1)$
environment of max or min value	any	$\frac{-b_1}{2b_2}$	$-\frac{b_1}{2b_2} - b_1\frac{\sigma^2(g_2)}{b_2^3} + \frac{\sigma(g_1,g_2)}{2b_2^2}$	$\frac{b_2^2\sigma^2(g_1) + b_1^2\sigma^2(g_2) - 2b_1 b_2\sigma(g_1,g_2)}{4b_2^4}$
max or min value	any	$a - \frac{b_1^2}{4b_2}$	$a - \frac{b_1^2}{4b_2} + \frac{b_2^2\sigma^2(g_1) + b_1^2\sigma^2(g_2) - 2b_1 b_2\sigma(g_1,g_2)}{4b_2^3}$	$\frac{b_1^4\sigma^2(g_2) + 4b_2\left(b_1^3(-\sigma(g_1,g_2)) + b_1^2 b_2\left(2\sigma(f,g_2) + \sigma^2(g_1)\right) - 4b_1 b_2^2\sigma(f,g_1) + 4b_2^3\sigma^2(f)\right)}{16b_2^4}$

704

Figure 1: Polynomial approximations to a non-linear reaction norm. The black curve represents an hypothetical true reaction norm of the form $E(z) = 1 + e^{0.75x - 0.15x^2 - 0.04|x|^3}$, which has the basic shape often expected for a thermal performance reaction norm. The solid lines show the predictions of polynomial approximations of the reaction norm of first- (red), second- (blue) and third-order (orange). Panel (a) shows the polynomial approximations assuming that nine environment-specific population mean phenotypes are known with essentially no error (e.g., as though there were very high sample sizes). Panel (b) repeats the true (black line) and approximated (solid coloured lines) reaction norms from (a), and in addition shows three more approximations of the reaction norm, in dashed lines. These are the polynomial approximations to the reaction norm that would be obtained, given the same true reaction norm, but if the environmental covariate was normally-distributed with mean zero, and with the same variance as among the nine treatments in part (a).
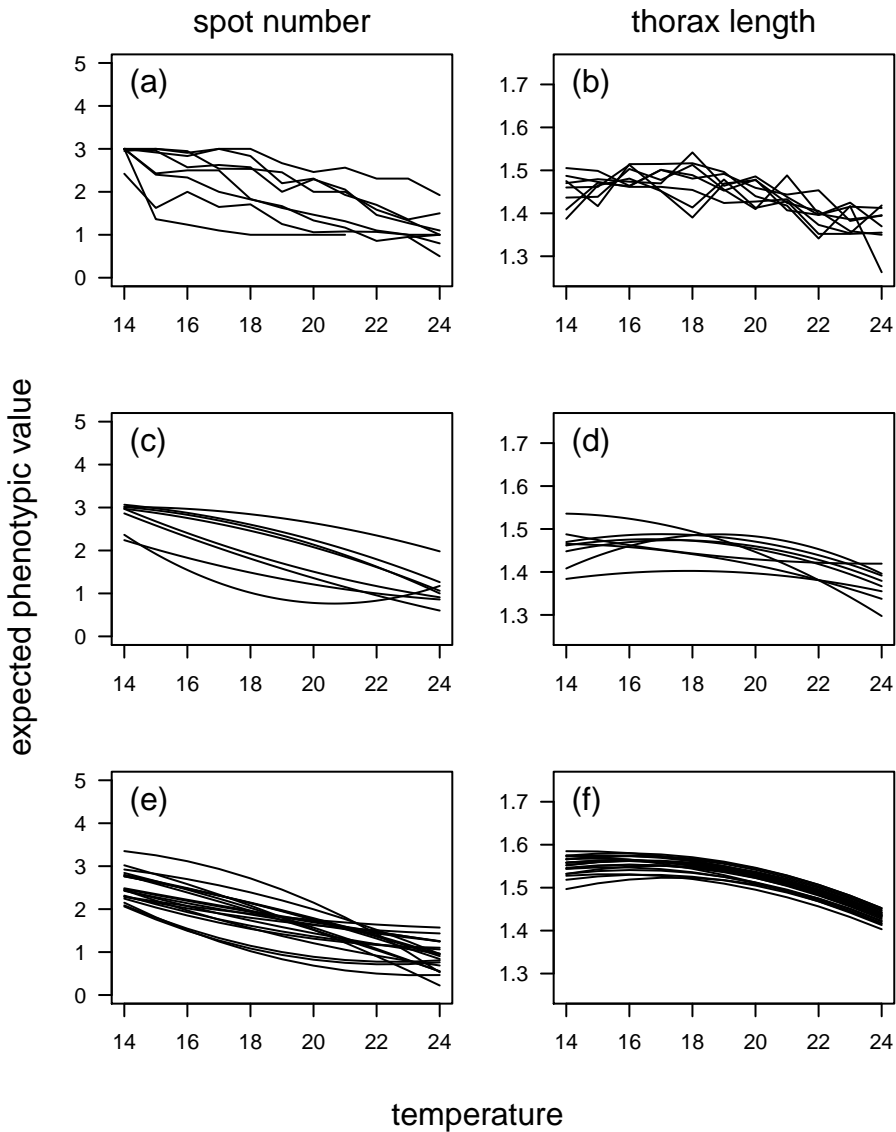
Figure 2: Representations of variation in reaction norm shape for abdominal spot number (left column) and thorax length (right column) among eight strains of *Drosophila mediopunctata* (raw data from Rocha et al. 2009). (a) and (b) show strain- and temperature-specific means, (c) and (d) show quadratic approximations to the strain-specific mean temperatures. (e) and (f) show families of 20 simulated reaction norms from quadratic random regression mixed models fitted to the individual-based (rather than line mean) data.
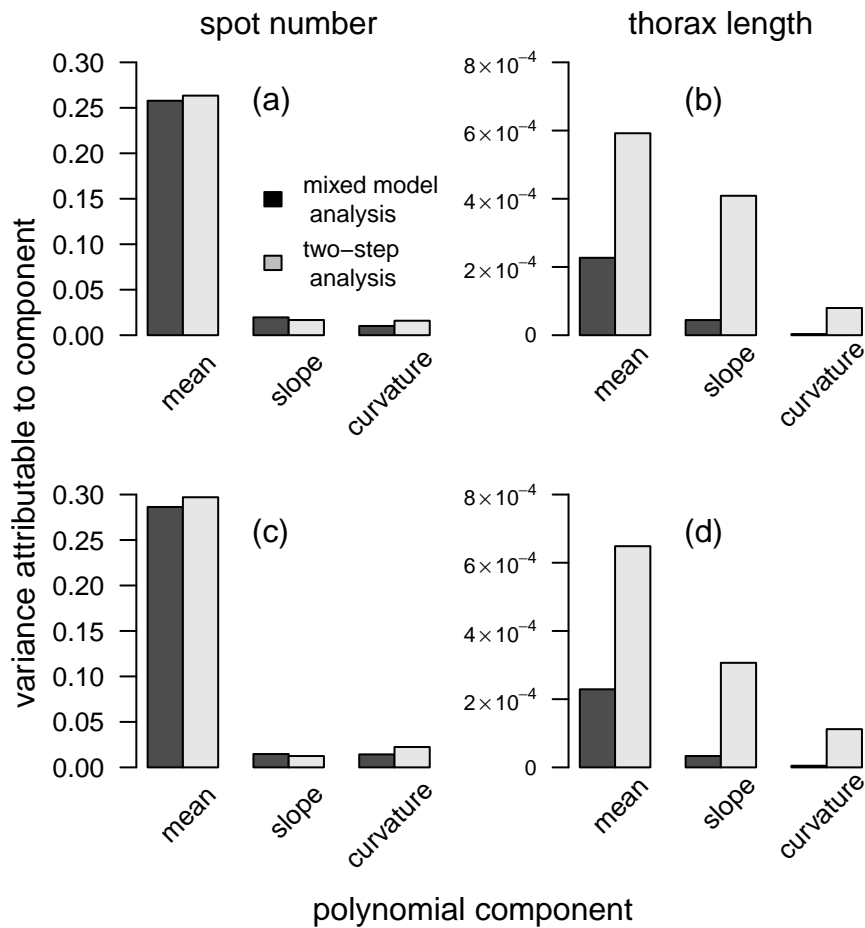
Figure 3: Proportions of variation around the average thermal reaction norms of abdominal spot number and thorax length in strains of *Drosophila mediopunctata* (raw data from Rocha et al. 2009), attributable to variation in overall means, slopes, and intercepts. (a) and (b) are calculated for a uniform covariate from 14 to 24 degrees (i.e., $r = 5$). (c) and (d) are calculated for a normal covariate centred on 19 degrees, and with a standard deviation of $r/2$, such that approximately 95 percent of the values of the environmental covariate would fall in the range investigated (i.e., between 14 and 24). Black bars show variances attributable to the polynomial components derived from a mixed model analysis (as depicted in figure 4e,f), and grey bars show those from the two-step analytical procedure.

Box 1: Mean centring of covariates

In linear regression analyses of reaction norms, one may be interested in how much mean values vary, and how much slopes vary (and perhaps how means and slopes covary). The intercept only represents the mean if the environmental covariate is mean-centred. Figure B1.1 illustrates why this is so: if functions are sloped, then intercepts can be very different from mean values. If two regression functions have different slopes, then differences in their intercepts may be very different from differences in their mean values, if the covariate is not mean-centred (figure B1.1a). Intercepts do represent the means for a centred covariate, regardless of how slopes differ among genetic units (figure B1.1a). Furthermore, mean-centring can alleviate artifactual correlations among parameters. While there is no correlation of means and slopes in the reaction norms depicted in figure B1.1 parts a and b – in fact there is no variation in means – slopes and intercepts are highly correlated in part a, but the true biological pattern of no covariance is reflected under mean-centring, as depicted in b.

Similarly to how intercepts cannot be interpreted as mean values of a regression function when a covariate is not centred, linear terms may be unrelated to average slopes of quadratic regression functions, unless the covariate is mean-centred (and symmetric). Linear terms in quadratic regression models are the slope of the function at the point where the covariate is zero. Consider figure B1.1c: the two depicted lines have identical average slopes over the range of the environment from three to five, and identical slopes at $x = 4$. However, if three to five environmental units is the range over which data have been collected, these two regressions will have opposite linear terms in a regression analysis where the covariate is not centred. Figure B1.1d depicts regression analyses of the same data, but with a mean-centred covariate, where it is evident that the linear terms will have the same value, reflecting the fact that the two approximating functions have the same slope at the mean value of the covariate.
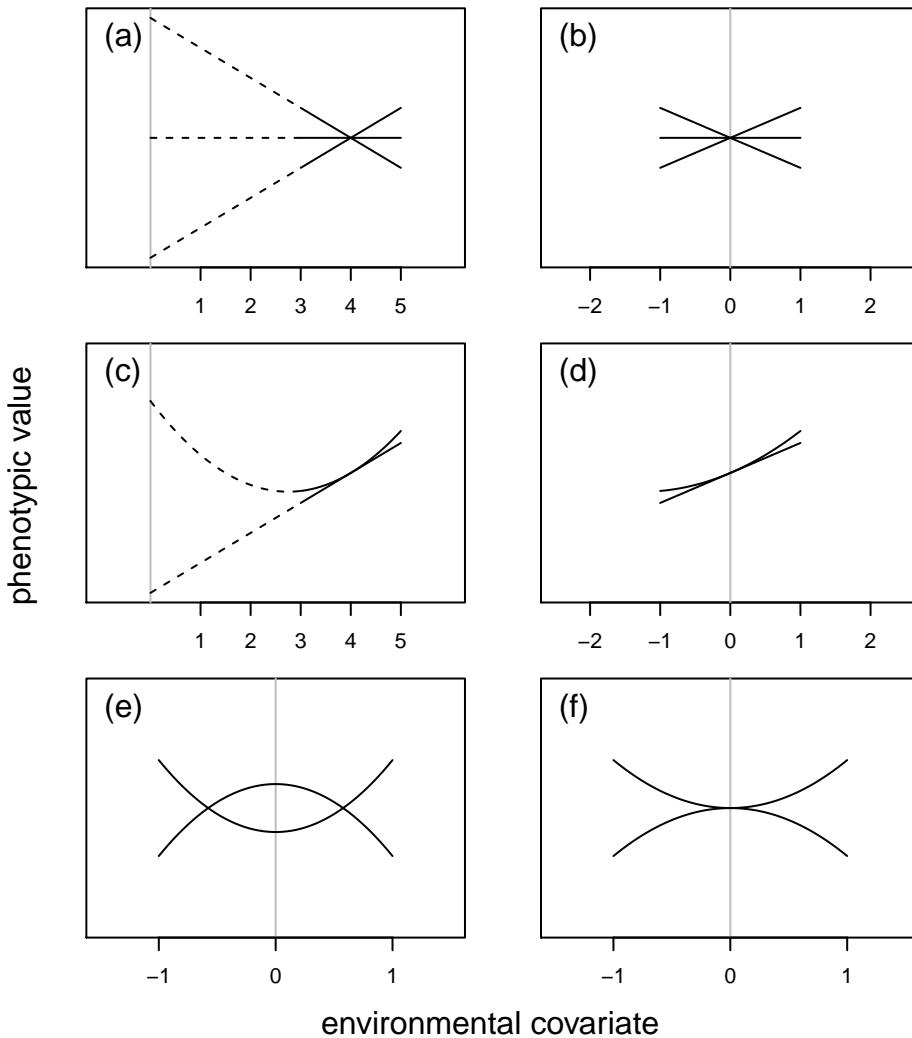
Figure B1.1: Illustrations of conditions under which mean-centring of an environmental covariate can, and cannot, render regression intercepts interpretable as mean values. See text in Box 1 for details.

730   In quadratic regression analyses of reaction norms, intercepts do not reflect mean values,

731 even under mean-centring, unless there is no curvature. If curvature varies among genetic

732 units, then then differences in intercepts may need not represent differences mean values.

733 Figure B1.1e,f illustrate two alternatives: in part e, intercepts vary between two reaction

734 norms, but mean values do not. In part f, intercepts are identical but mean values differ.

735 This does not mean that quadratic regression analysis of reaction norms cannot provide

736 inference of variation in mean values, only that these values must be derived (see table 3).

Box 2: Character-state and reaction norm approaches: the maths

Any family of reaction norms, summarised by their mean intercept and slope, and the variances and covariances of slopes and intercepts, corresponds to a specific set of environment-specific means and variances of phenotype, and across-environment covariances of phenotype. Expressing a family of reaction norms (or of function-valued phenotypes generally, e.g., developmental trajectories) in terms of environment-specific (or, for e.g., age-specific in the case of development) means and (co)variances is referred to as a character-state representation. Any complete description of the covariances of parameters of a family of polynomial reaction norms can be translated into what it implies about a character-state representation. The maths involved may initially seem nebulous, but are in fact reasonably straight-forward.

The algebraic operation necessary to convert a characterisation of reaction norms to a character state representation is the expression for the variance of a random variable when subjected to a linear transformation. If x is related to y according to $y = bx$, then the variance of x is related to the variance of y according to $\sigma^2(y) = b^2\sigma^2(x)$. If $\mathbf{x}$ and $\mathbf{y}$ are vectors, then a linear transformation might be written $\mathbf{y} = \mathbf{Bx}$, where $\mathbf{B}$ is a matrix containing coefficients by which elements of $\mathbf{x}$ are related to elements of $\mathbf{y}$. If instances of $\mathbf{x}$ have (co)variances $\mathbf{\Sigma_x}$, then the (co)variances of $\mathbf{y}$s are given by $\mathbf{\Sigma_y} = \mathbf{B\Sigma_x B}^T$, where $\mathbf{B}^T$ is the transpose of $\mathbf{B}$ (i.e., a matrix where rows and columns are exchanged).

To use the algebra of variances under linear transformation to convert covariances of polynomial coefficients of reaction norms into their character-state representation, we must compose matrices $\mathbf{B}$ that reflect the environments in which we want to express the variances, and among which we might want to know covariances. An example might be most useful at this point. The estimated covariances of intercepts, slopes, and quadratic terms of reaction

760 norms of spot number to temperature are

$$\Sigma_x = \begin{bmatrix} 0.381 & \text{-}7.89^{-3} & \text{-}8.18^{-3} \\ \text{-}7.89^{-3} & 2.35^{-3} & 2.97^{-4} \\ \text{-}8.18^{-3} & 2.97^{-4} & 1.83^{-4} \end{bmatrix}.$$

761   The quantities by which reaction norm parameters must be multiplied to give expected

762 phenotype in any given environment are polynomial values corresponding to that environ-

763 ment's numerical value. In the analyses of *Drosophila* reaction norms, the temperature data

764 were centred to a mean of 19°C. So, the values of the polynomial function for, say, 16, 19 and

765 22°C (the middle value and nearly the extremes of the temperature range, see figure 2), the

766 polynomials would be $[(16 - 19)^0 \ (16 - 19)^1 \ (16 - 19)^2]$, $[(19 - 19)^0 \ (19 - 19)^1 \ (19 - 19)^2]$

767 and $[(22 - 19)^0 \ (22 - 19)^1 \ (22 - 19)^2]$. The variances and covariances of the reaction norm

768 functions across the temperatures 16, 19 and 22°C would then be given by

$$\begin{bmatrix} 1 & -3 & 9 \\ 1 & 0 & 0 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} 0.381 & \text{-}7.89^{-3} & \text{-}8.18^{-3} \\ \text{-}7.89^{-3} & 2.35^{-3} & 2.97^{-4} \\ \text{-}8.18^{-3} & 2.97^{-4} & 1.83^{-4} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -3 & 0 & 3 \\ 9 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 0.301 & 0.331 & 0.228 \\ 0.331 & 0.381 & 0.284 \\ 0.228 & 0.284 & 0.239 \end{bmatrix}.$$

769   This covariance matrix represents the covariances at the genetic level, in this case among-

770 strains, at which reaction norms are inferred. The high correlations among environments

771 (covariances are positive and similar in magnitude to variances) reflect the character-state of

772 representation of the fact that there is modest variation in slopes and curvatures of reaction

773 norms.

774   In general, a reaction norm approach will use fewer parameters than a character state

775 approach. In the *Drosophila* example, a description of the strain-level covariance matrix of

776 temperature specific phenotype would require estimation of a matrix with 66 parameters.

777 In this specific example with eight strains, these parameters could not be simultaneously

778 estimated. In fact, any pairwise covariance estimate should be regarded as tenuous given

779 this level of replication across strains. In contrast, the quadratic reaction norm approach

780 estimates six parameters at the level of strain. One should still keep in mind that its inferences

781 are based the only eight strains, but all the information available is simultaneously used to

782 estimate a model with a more sensible number of parameters. In cases where there is sufficient

783 replication to support both character state and reaction norm approaches, their comparison

784 should be useful. For example, such comparisons could identify ranges of the environmental

785 covariate where a low-dimensional random regression model fits adequately or otherwise.

786 For smaller studies, visual comparison of random regression fits to raw data is probably best

787 (figure 2).

788

789 _____