Variation in reaction norms: statistical considerations and biological interpretation

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

June 15, 2016

[1]School of Biology, University of St Andrews

[2]Department of Animal Ecology, VU University Amsterdam

[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

Abstract

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

Introduction

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

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

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

of reaction norms. As examples of such reaction norm properties, we consider inference both of properties of reaction norms of individual genotypes (or other genetic groupings such as populations; e.g., focusing on their slopes, or the locations of their maxima), and properties of families of reaction norms (e.g., variation in slopes, or variation in the locations of maxima). The primary focus is on biological inference, but we draw extensively on the statistical theory underlying different potential analytical approaches to studying reaction norms. In some cases, we expand basic theory about regression analysis to yield new insights about how specific, biologically-motivated, regression analyses may behave. We discuss biological inference of properties of reaction norms in general, but we also specifically focus on on some recent claims that have been made about the efficacy of different approaches.

Polynomial regression, and especially quadratic regression, is potentially very useful for characterising reaction norms, and several authors have investigated theoretical and empirical properties of reaction norms using such functions as theoretical and statistical models (e.g., Delpuech et al. 1995; Gavrilets and Scheiner 1993). Two recent very firm claims about analysis of reaction norms with polynomial functions are: (1) that the slope of a line connecting mean phenotype in two environments is generally misleading about the form of a reaction norm (Rocha and Klaczko, 2012); and (2) that quantities derivable from polynomial regressions, such as the slope at any point, or measures of overall curvature, provide robust inference of reaction norms (Rocha and Klaczko, 2014). We show analytically, and with numerical examples, that neither of these assertions is generally true. Nonetheless, we agree that polynomial regression, perhaps especially quadratic regression, may be very useful for biological studies of reaction norms. However, polynomial regression will be most useful if applied with a somewhat more nuanced understanding of its strengths and limitations.

We also contrast two approaches to characterising variation in reaction norms. By "characterising variation", we refer to situations where we are not necessarily interested in specific reaction norms, nor in comparisons of properties of two or few specific reaction norms, but rather where we seek to assess variation in populations for aspects of reaction norms. For

example, we may be interested in how much variation in average slope, relative to variation in mean values, occurs among the reaction norms of the genotypes segregating within some population, or among populations within a species. There are two basic approaches in use to quantifying such variation. In the first procedure, two steps are employed. First, data from each genetic group (individual, genotype, inbred line, etc.) are subjected to statistical analysis, for example, to regression analyses to determine slopes, as well as calculations of line-specific means across environments. In the second step of the first approach, summary statistics are calculated at the population level, providing, for example, measures of variance in the means and slopes estimated in the first step. In the second type of approach, mixed models, in particular, random regression mixed models, may be used to directly estimate variance in reaction norm parameters. We show analytically how the two-step approach introduces biases into most inferences about variation in reaction norms, and we illustrate the application of random regression mixed models, in detail, with an empirical example. We also derive new measures of variation in phenotype arising from different aspects of reaction norms, and show how these may be particularly useful for answering questions of current interest in reaction norm research.

This paper is arrayed in several sections. In each, biologically-relevant results and the more intuitive pieces of statistical theory upon which they rest are presented, while more involved statistical theory is generally relegated to an extensive appendix. First, we consider methods for characterising aspects of individual reaction norms, including the slope between mean phenotype in two environments, and polynomial regression. We then turn our attention to inference of variation in reaction norms. We compare the two-step and mixed model-based approaches, present an empirical example, and derive new measures of variation in reaction norm shape. In the discussion, we recapitulate our major points and address various common threads.

Slopes between two points

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

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

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

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

Simplifying this expression using basic algebra and calculus rules gives

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

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

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

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

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

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

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

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

with respect to minimising the sampling variance of the average slope is demonstrated more rigorously in appendix section A.1.

Polynomial regression analysis of reaction norms

Polynomial regressions are proven statistical tools for characterising functions, and have been advocated for analysis of non-linear reaction norms (e.g., Gavrilets and Scheiner 1993; Rocha and Klaczko 2014). Polynomial regressions will typically be least-squares fits of an approximating function to a true reaction norm with an unknown true functional form. Given phenotypic values, z, and environmental values, x, for units of observation indexed i, first- (linear), second- (quadratic), and third-order (cubic) polynomial regressions take the form

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

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

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

In each fitted regression model, the intercept, a, and polynomial regression coefficients, i.e., b_1, b_2 and b_3, will be those that minimise the variance of the residuals (e). Note that the values of the intercept and common coefficients (e.g., b_1) may differ between models of different polynomial order, fitted to the same data.

In application of polynomial regression, it is hoped that coefficients of the regression model, or predictions from the fitted model, will reflect biologically relevant aspects of reaction norms. While polynomial regression may often be pragmatic, the conditions under which coefficients of polynomial regression models will reflect specific, biologically relevant, quantities such as the average slope of curvature are limited. Where there is a very simple and general interpretation of s_{ab} as the average slope of a true arbitrary reaction norm between the points $x = a$ and $x = b$, the conditions under which a polynomial regression can

provide a similar inference are much more limited. If the environmental variable is normally distributed - both in the data analysed and in the relevant scenario in nature about which we want to draw inferences, then the linear term (b_1) in a first-order quadratic regression model gives the average slope. This can be demonstrated from Stein's lemma (Stein, 1973), 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. A regression coefficient is the covariance of the predictor and the response, divided by the 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 regression models estimates the average slope as well, still requiring normality of the environmental variable, and also that the environmental variable has a mean of zero. Similarly, if the environmental covariate is normally-distributed and mean-centred, the quadratic term in a quadratic approximation to the reaction norm is equal to half of the average second derivative of the reaction norm function. These specific properties of quadratic regression analysis, when covariates are normal and mean-centred, underlie regression-based analysis of selection gradients, which are the average first and second (partial) derivatives of (relative fitness) functions as well (Geyer and Shaw, 2010; Lande and Arnold, 1983; Mitchell-Olds and Shaw, 1987).

So, there is a condition, namely, normality of the environmental covariate x, under which coefficients of polynomial regressions have very general and biologically useful interpretations. Under normality, the slope and curvature of a polynomial approximation to a reaction norm can reflect the average slope and curvature of the true reaction norm, regardless of the true form of the reaction norm. However, the condition of normality of the environmental variable (x) is patently not met in virtually all studies of reaction norms. Rather, by design, the distribution of the environmental variable(s) are non-normal, being composed of two or more discrete treatments. This distribution for x tends toward a uniform distribution as the number of treatments increases. If the distribution of the environmental variable is not normal, then the parameters of a polynomial regression have no direct biological interpretation, and no single useful statistical interpretation, other than that they are the parameters that minimise the residual variance. However, parameters of a polynomial regression will provide

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

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

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

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

Assessing variation in reaction norms

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

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

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

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

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

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

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

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

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

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

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

ation in two-step analytical procedures, and (2) that the extent to which statistical noise pollutes biological inferences depends on scaling, are not artefacts of the simplifying assumptions made here. For example, it occurs if the covariate is not strictly uniform, but rather is composed of few or many distinct environmental treatments. The pattern will also hold for very different distributions of the covariate; analogous expressions for $\Sigma[\mu]$, $\Sigma[b_1]$, and $\Sigma[b_2]$, given a normal covariate, x, are given in appendix section A.5.

A class of linear mixed models called random regression models exists specifically to separate noise from real variation in families of regression coefficients. The simplest random regression mixed model is a linear random slopes model, which can be written as

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

where z_{ij} is the phenotypic observation of individual i from group (e.g., species, family, etc.) j, where a and b are fixed regression parameters for the intercept and slopes, respectively, where f_j and g_j are regression parameters (contrasts to a and b) for group j, x_i is the environment to which individual i was exposed and e_i is a residual for individual i. As before, the residuals are assumed to be drawn from a normal distribution $e_i \sim N(0, \sigma_r^2)$, and furthermore, the group-specific regression parameters are also treated as random variables, i.e., variables that belong to a bivariate normal distribution

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

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 and intercepts. Solutions to the mixed model give estimates of terms including the variance in slopes ($\sigma^2(g)$), that are not inflated by sampling error, as occurs in the two-step approach. While random regression analysis is currently in use (Dingemanse et al., 2010; Martin et al., 2011), we hope that it is useful to clarify that its use represents more than a mere modernisa-

tion of statistical approaches to studying reaction norms. Random regression can yield direct inferences of variation in reaction norm parameters (e.g., of slopes), that would otherwise be subjected to potentially biologically misleading statistical biases in two-step analyses.

A random regression model can be used to assess variation in reaction norms for any analysis with multiple units of observation, and two or more (or a continuous range of) environmental treatments. When applied to a study with two treatments, the linear random regression mixed model specified by equation 5 yields unbiased estimates of the among-group variance in intercepts, average slopes (i.e., this amounts to an analysis of variation in s_{ab}), and their covariance.

When applied to a study with a range of environmental conditions, or with random quadratic (or even higher order) terms, random regression mixed model analysis can be used to recover meaningful information about variation in reaction norm shape. However the caveats that apply to the interpretation of polynomial approximations to reaction norms in general will also apply to inferences about variation in polynomial coefficients obtained by random regression. With prudence, it is possible that random polynomial regression mixed model analysis could be much more extensively used in analysis of variation in reaction norms, and such analysis will certainly be preferable to two-step analytical approaches in most circumstances.

Example application of a random regression mixed model

We applied quadratic random regression mixed model analysis to the data on reaction norms reported in Rocha et al. (2009) and re-analysed in Rocha and Klaczko (2012) (data provided by F. B. Rocha and L. B. Klaczko). The data consist of 1122 *Drosophila mediopunctata* phenotyped for abdominal spot number and thorax length, raised in three simultaneous replicates (vials) in a thermal gradient spanning 14°C to 24°C in 1°C intervals. For each

trait, the mixed model took the form

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

where $z_{i,j}$ represents the phenotype (spots or thorax length) measured on individual i belonging to strain j. The fixed effects, a, b_1, and b_2 estimate the average reaction norm, conditional on a fixed effect of sex. The random polynomial coefficients f_j, $g_{1,j}$, and $g_{2,j}$ for each line and are assumed to be drawn from a multivariate normal distribution

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

with estimated covariance matrix $\mathbf{\Sigma}$. Additionally, the replicate associated with individual i, and the temperature in which it was raised, coded as a multi-level factor t_i and the residuals, e_i, are all included as random effects with estimated variances. The temperature at which a given individual was raised, t_i was mean-centred by subtracting 19°C.

The among-line covariance matrices of intercepts, slopes, and curvatures (table 1) are difficult to interpret directly. However, some features of the mixed model analysis are immediately apparent. First, we can see that, as predicted by the statistical theory given above, the variance of coefficients of reaction norms (table 1d) in the two-step procedure inflates the apparent amount of variation in reaction norm parameters. Another such comparison yielding similar inflation of apparent variation in reaction norm parameters is reported in Liefting et al. (2015). This effect is larger in cases when reaction norms are more similar, in this case, with a much more dramatic effect for thorax length reaction norms than for spot number. Furthermore, the correlations among reaction norm parameters are consistently smaller in inferences from the two-step procedure. This is because statistical noise inflates all of the estimates of variance in the polynomial coefficients, but not necessarily all of the covariances.

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

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

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

2b; if all of the reaction norms were nearly identical, then statistical noise in estimating the mean phenotype for each strain in each environment would cause half of the line segments in a plot such as this to cross. Clearly, line-crossing is difficult to apply as a measure of reaction norm complexity.

Variance in phenotype arising from variation in reaction norm parameters

How can we make inferences about the relative importances of variation in the mean values of reaction norms, and of slopes and quadratic terms, if the variances of intercepts, slopes, and quadratic terms depend on the arbitrary scaling and distribution of the environmental covariate (and additionally on the covariance of intercepts and quadratic terms)? Given any distribution of the environmental covariate, it is possible to derive the amount of variation among genetic units (e.g., lines), integrated over the distribution of the covariate, that arises from differences in the mean, slope and quadratic curvature of reaction norms. Detailed derivations are given in the appendix (section A.6 for a uniform covariate and A.7 for a normal covariate), and the formulae for these measures of variation in different aspects of reaction norm shape are given in table 2.

Figure 3 shows the amount of variation in expected line- and temperature-specific phenotypic values for both traits in the *Drosophila* example that are attributable to variation in the means, slopes, and curvatures of the families of reaction norms, as assessed by the quadratic random regression mixed model analysis (equation 6), and by the two-step analytical procedure. This quantification of different components of variation in reaction norms confirms that differences in reaction norm shape are indeed modest, and also further demonstrates the danger of inflating inferences in the two-step analytical procedure (figure 3a,b). The majority of variation among lines arises from differences in mean values of reaction norms. Calculations of variation attributable to intercepts, linear, and quadratic terms, separately made based on uniform and normal distributions, are quite similar in this example (compare figures 3a,b with 3c,d). This should generally be the case when families of quadratic functions

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

(Co)variances of arbitrary reaction norm properties

Intercepts, linear, and quadratic terms, and the variance in expected values with which they are associated, do not directly represent all features of reaction norms in which we may be interested. For example, we may be interested in phenotypic values at specific environmental values, mean phenotypic values integrated over different distributions of environmental values, locations of maxima or minima (environments that produce minimum or maximum phenotypes), and phenotypic values at maxima or minima (minimum or maximum phenotypic values). Quadratic regressions contain information about such reaction norm properties, previously sometimes termed "characteristic values" (Delpuech et al., 1995; Gibert et al., 1998), and families of regression coefficients, as estimable by random regression analysis, contain information about means and variances of such reaction norm properties. Operationally, calculations of variance in arbitrary reaction norm properties seems easiest in the two-step analytical procedure. In mixed model analyses, one must call on somewhat more statistical sophistication to derive (co)variances of reaction norm properties from estimated means, variances, and covariances of quadratic reaction norm parameters. However, an approach to develop formulae for such quantities seems clear. Given a function for calculating some quantity (e.g., the location of an optimum) from a fitted reaction norm function, the variance of that quantity can be approximated by taking the expectation of a Taylor series. It is reasonable to feel that this is easier said than done. It may therefore be useful to provide expressions involving some quantities that might be most useful.

Table 3 gives expressions for quantities that may be calculated from quadratic reaction norm approximations: the environment of the maximum or minimum value, the maximum or minimum value, and the mean (remembering that the intercept is not the mean of a quadratic function, even if the covariate is mean centred) for different distributions of the environmental covariate. Table 3 also gives expressions for the expectations and variances of each of these

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

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

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

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

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

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

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

Discussion and Conclusions

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

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

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

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

features of (variation in) reaction norms. However, the analyses here are heartening and it seems plausible, that with due consideration to the features of any particular study system, that quadratic regressions, as advocated by (Rocha and Klaczko, 2014), could indeed provide pragmatic models of reaction norms in many cases.

However, we caution strongly against some of Rocha and Klazcko's (2014) specific suggestions for interpreting quadratic regressions. In particular, Rocha and Klazcko suggest that the derivative of a quadratic, or other polynomial function at any specific point, which they call "local reaction norm plasticity" could be a generally useful measure of reaction norm shape at a particular point. However, this derivative need not necessarily closely reflect the slope of a true reaction norm at that point, and it need not even be the correct sign (see also appendix section 2, figures A.1 and A.2). Rocha and Klaczko (2014) also suggest that the quadratic term can be used as a measure of "reaction norm shape", justified by the fact that twice the value of the quadratic term is the second derivative of the quadratic function at all points, and is therefore the average derivative of the quadratic function. This use seems reasonable, but its application should be approached with awareness that the average derivative or second derivative of the quadratic approximation to any function is only equal to the average derivative or second derivative under two conditions. First, this equality holds if the true function is indeed quadratic. Second, this equality holds if the environmental covariate is normal. Virtually no studies of reaction norms have a normal covariate. In fact, investigators typically strive for covariate distributions that approach uniformity. So justification for using curvature of a quadratic approximation as a measure of the curvature of a reaction norm rests on a requirement that a quadratic function is a good approximation of the true reaction norm. It seems that this requirement should frequently be closely enough met in empirical systems for quadratic regression to provide useful measures of average reaction norm curvature.

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

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

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

Acknowledgements

F. B. Rocha and L. B. Klaczko provided the *Drosophila* data. Mike Ritchie and Kerry Johnson provided comments on earlier drafts of this paper, which was also greatly improved by comments from two anonymous reviewers. MBM is supported by a Royal Society (London) University Research Fellowship. ML is supported by the Netherlands Organisation for Scientific Research, VIDI grant nr. 864.03.003.

References

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

Berg, M. P., E. T. Kiers, G. Driessen, M. van der Heijden, B. W. Kooi, F. Kuenen, M. Liefting, H. A. Verhoef, and J. Ellers. 2010. Adapt or disperse: understanding species persistence in a changing world. Global Change Biology 16:587–598.

Chevin, L. M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLOS Biology 8:e1000357.

Delpuech, J.-. M., B. Moreteau, J. Chiche, E. Pla, J. Vouidibio, and J. R. David. 1995. Phenotypic plasticity and reaction norms in temperate and tropical populations of *Drosophila melanogaster*: ovarian size and developmental temperature. Evolution 49:670–675.

DeWitt, T. J. and S. M. Scheiner, 2004. Phenotypic plasticity - Functional and Conceptual Approaches. Oxford University Press, Oxford.

Dingemanse, N. J., A. J. Kazeem, K. Réale, and J. Wright. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends in Ecology and Evolution 25:81–89.

Ellers, J. and G. Driessen. 2011. Genetic correlation between temperature-induced plasticity of life-history traits in a soil arthropod. Evolutionary Ecology 25:473–484.

Fallis, L., J. Fanara, and T. Morgan. 2014. Developmental thermal plasticity among *Drosophila melanogaster* populations. Journal of Evolutionary Biology 27:557–564.

Gavrilets, S. and S. M. Scheiner. 1993. The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. Journal of Evolutionary Biology 6:31–48.

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

Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394–407.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rocha, R. B., H. F. Medeiros, and L. B. Klaczko. 2009. The reaction norm for abdominal pigmentation and its curve in *Drosophila mediopunctata* depend on the mean phenotypic 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})	$\frac{4}{45}r^4\sigma^2(g_2)$	$2\sigma^4(x)\sigma^2(g_2)$

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	$\frac{-b_1}{2b_2}$	$\frac{-b_2}{2b_2} - b_1\frac{\sigma^2(g_2)}{b_2^3} + \frac{\sigma(g_1,g_2)}{2b_2^2}$	$\frac{b_2^2\sigma^2(g_1)+b_1^2\sigma^2(g_2)-2b_1b_2\sigma(g_1,g_2)}{4b_2^4}$
max or min value	any	$a - \frac{b_1^2}{4b_2}$	$a - \frac{b_1^2}{4b_2} + \frac{b_2^2\sigma^2(g_1)+b_1^2\sigma^2(g_2)-2b_1b_2\sigma(g_1,g_2)}{4b_2^3}$	$\frac{b_1^4\sigma^2(g_2)+4b_2\left(b_1^3(-\sigma(g_1,g_2))+b_1^2b_2(2\sigma(f,g_2)+\sigma^2(g_1))-4b_1b_2^2\sigma(f,g_1)+4b_2^3\sigma^2(f)\right)}{16b_2^4}$

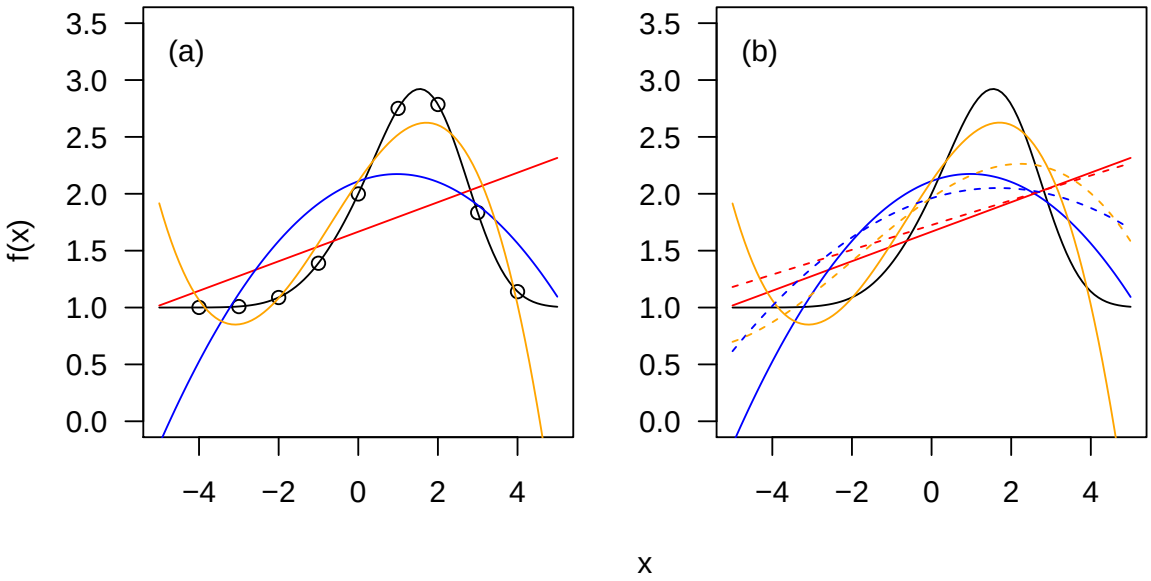

Figure 1: Polynomial approximations to a non-linear reaction norm. The black curve represents an hypothetical true reaction norm of the form $E(z) = 1 + e^{0.75x - 0.15x^2 - 0.04|x|^3}$, which has the basic shape often expected for a thermal performance reaction norm. The solid lines show the predictions 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).

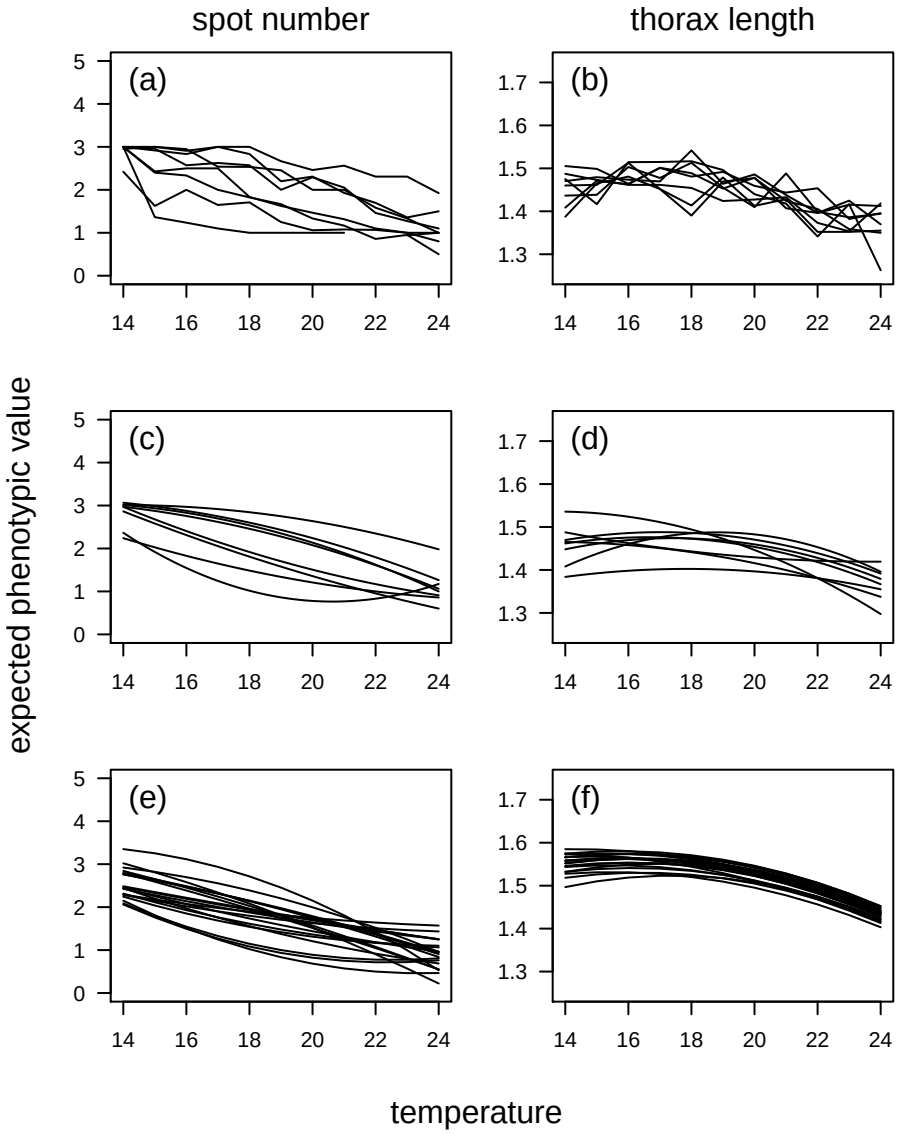

Figure 2: Representations of variation in reaction norm shape for abdominal spot number (left column) and thorax length (right column) among eight strains of *Drosophila mediopunctata* (raw data from Rocha et al. 2009). (a) and (b) show strain- and temperature-specific means, (c) and (d) show quadratic 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.

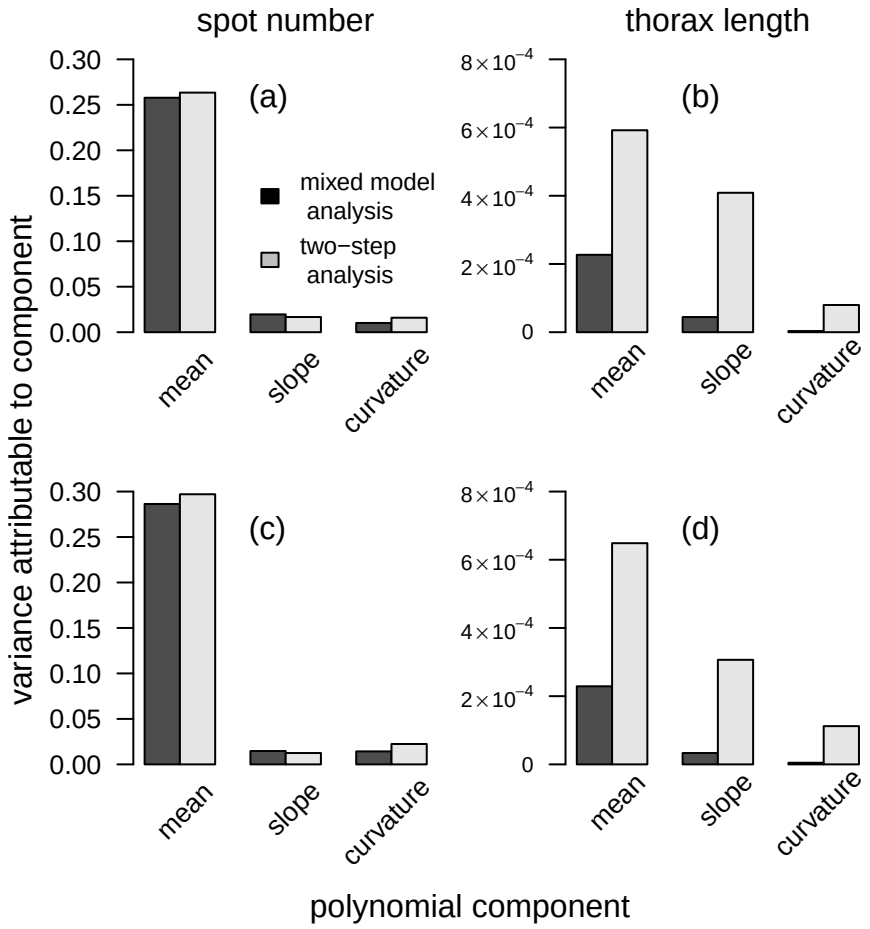

Figure 3: Proportions of variation around the average thermal reaction norms of abdominal spot number and thorax length in strains of *Drosophila mediopunctata* (raw data from Rocha et al. 2009), attributable to variation in overall means, slopes, and intercepts. (a) and (b) are calculated for a uniform 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.

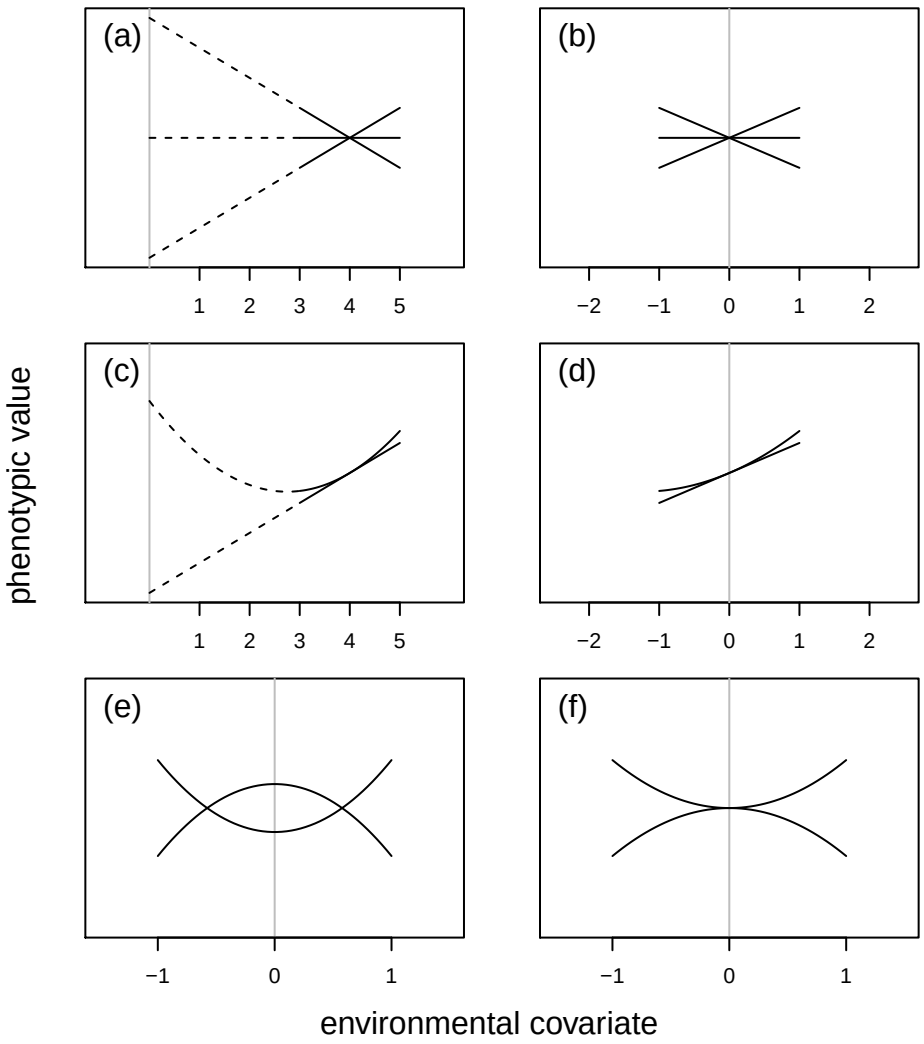

Figure B1.1: Illustrations of conditions under which mean-centring of an environmental covariate can, and cannot, render regression intercepts interpretable as mean values. See text in Box 1 for details.

In quadratic regression analyses of reaction norms, intercepts do not reflect mean values, even under mean-centring, unless there is no curvature. If curvature varies among genetic units, then then differences in intercepts may need not represent differences mean values. Figure B1.1e,f illustrate two alternatives: in part e, intercepts vary between two reaction norms, but mean values do not. In part f, intercepts are identical but mean values differ. This does not mean that quadratic regression analysis of reaction norms cannot provide 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 $\mathbf{y} = \mathbf{B}\mathbf{x}$, 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 $\mathbf{\Sigma_y} = \mathbf{B}\mathbf{\Sigma_x}\mathbf{B}^T$, 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

norms of spot number to temperature are

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

The quantities by which reaction norm parameters must be multiplied to give expected phenotype in any given environment are polynomial values corresponding to that environment's numerical value. In the analyses of *Drosophila* reaction norms, the temperature data were centred to a mean of 19°C. So, the values of the polynomial function for, say, 16, 19 and 22°C (the middle value and nearly the extremes of the temperature range, see figure 2), the polynomials would be $[(16-19)^0\ (16-19)^1\ (16-19)^2]$, $[(19-19)^0\ (19-19)^1\ (19-19)^2]$ and $[(22-19)^0\ (22-19)^1\ (22-19)^2]$. The variances and covariances of the reaction norm functions across the temperatures 16, 19 and 22°C would then be given by

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

This covariance matrix represents the covariances at the genetic level, in this case among-strains, at which reaction norms are inferred. The high correlations among environments (covariances are positive and similar in magnitude to variances) reflect the character-state of representation of the fact that there is modest variation in slopes and curvatures of reaction norms.

In general, a reaction norm approach will use fewer parameters than a character state approach. In the *Drosophila* example, a description of the strain-level covariance matrix of temperature specific phenotype would require estimation of a matrix with 66 parameters. In this specific example with eight strains, these parameters could not be simultaneously 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 ---