Review of *'Partitioning the Phenotypic Variance of Reaction Norms'.*

Jarrod Hadfield*

In this paper the authors detail a new way of decomposing phenotypic variance into parts due to i) average plasticity, ii) genetic variation in reaction norms and iii) residual variation. As outlined in the introduction, a major motivation for this work is to develop metrics for comparing plasticity across traits. However, some form of standardisation is still required to achieve this as the metrics are in units of traits squared. Some discussion of alternative standardisations (for example, dividing by the mean squared or total variance) should be made. This issue aside, I had mixed feelings about the novelty/utility of the approach. Component i) (V_{Plas}) is straightforward and sensible yet it is essentially an (adjusted) R^2 from the relevant fixed effect part of the model and so doesn't, in my opinion, require a lengthy paper discussing it. Similarly, metrics to distinguish between a curve-parameter function and a character state approach seemed a little ad-hoc given well known tests such as likelihood-ratio tests or F-tests could be employed. Despite this lack of novelty, I do think the question is interesting, and I would advocate the approach if I was trying to characterise the environmental sensitivity of a trait.

When reading the introduction (for example, the main paragraph of P3) I thought the main innovation would be in developing methods for comparing genetic variation in plasticity. This is a difficult problem since plasticity is always in units that is a function of the trait and environment units, an issue the authors point out in the context of the study by Murren et al. (2014). However, component ii) (V_{Gen}) does not really characterise genetic variation in plasticity in my view: it is simply the average genetic variation within environments. The limitations of this are most obvious in the character state approach where the metric is completely insensitive to the genetic correlations between environments. Most people would consider the sign and magnitude of these correlations to completely represent the magnitude of genetic variation in plasticity (GxE). Attempts have been made to develop metrics for quantify GxE in discrete environments (most recently, Albecker et al. (2022)) and I wonder if these ideas could be fruitfully extended to reaction norm approaches?

- L56: It's not clear why standard statistical methods cannot be used to assess fit in this context (e.g. AIC or likelihood-ratio tests).

*I sign all reviews.

- L116: ϵ_k is not defined but presumably stands for a (single) environmental variable.
- L118-119 '*such approach*' should read '*such an approach*'.
- L124 '*function-values traits*' should read '*function-valued traits*'.
- L139 The law of total variance given is the one for conditioning on a single variable, not two. For two variables, it would be:

$$V(\hat{z}) = E[Var(\hat{z}|\epsilon, g)] + E[Var(E[\hat{z}|\epsilon, g)]|\epsilon) + Var(E[\hat{z}|\epsilon]) \quad (1)$$

  I guess the simplification to one variable comes about because $\hat{z}$, as opposed to z, is fully defined by ϵ and g such that $Var(\hat{z}|\epsilon, g) = 0$. Whether this is worth mentioning in the text, I'm not sure - it might unduly complicate the paper?
- L143 In the introduction it is pointed out that plasticity measures will have different units and scales depending on the trait and environment, and this makes it hard to make cross-trait comparisons. Based on the introduction, the reader is expecting this issue to be resolved, yet V_{Plas} is not unitless - it is in the units of the trait squared. What would the recommendation be? Standardise by the phenotypic variance or the mean or something else?
- L145 I think it would be easier to simply refer to V_{Gen} as the average genetic variance within environments throughout the manuscript. While I think $V_\epsilon(E_{g|\epsilon}(\hat{z}))$ has the straightforward meaning attached to it, I think $E_\epsilon(V_{g|\epsilon}(\hat{z}))$ does not, as is pointed out briefly in the paper. Imagine the case where there is no plasticity whatsoever, only genetic variation (in the intercept). Then, $E_\epsilon(V_{g|\epsilon}(\hat{z}))$ is simply the additive genetic variance in a standard non-plasticity model and it would be odd to refer to this as the genetic variance in reaction norms or genetic variance in plasticity. Consequently, I don't think this would be my choice of metric for comparing, say, levels of genetic variance for plasticity across traits. Similarly, imagine a discrete character state model - the metric is insensitive to the genetic correlations between environments (i.e. Equation 17), despite this being the major signal of genetic variance in plasticity (i.e. GxE). In addition, in most cases (where the average reaction norm is modelled using fixed effects) V_{Plas} is simply the variance explained by the terms, which when scaled by the total variance, would be the R^2 of the model. I think this simple fact will be lost on the less mathematical readers and should be stated.
- L184 This partition changes depending on arbitrary choices of what consitutes the reference environment, and in fact by chosing the reference environment to be the mean, this minimises the covariance and maximsises the variances (at least when ϵ is symetric, not sure otherwise). I think this is fine, but it would be good to explicitly state this.

- L196 *'includes all exponentiation levels (up to n) of the environmental variable ϵ'* is perhaps better stated as *'includes the environmental variable ϵ taken to all powers from 0 to n*

- L203 Sometimes orthogonal polynomials (e.g. Legendre) are used in statistical analyses. This would get rid of these issues - would the authors recommend them?

- L216 It may be worth noting that this is equivalent to Equation 11 in (Johnson 2014) although there the variance is conditional on observed values of ϵ rather than considering a probability distribution for ϵ. Note also, that Johnson's (2014) aim is to develop an R^2 metric for random-regression models. However, the result is inexact depending on how you want to define the R^2 because $E[V_{explained}]/E[V_{total}]$ (Johnson's (2014) metric where the expectation is over ϵ) is not equal to $E[V_{explained}/V_{total}]$ (see here).

- L266 Eq. 26 It would be nice to relate this to the adjusted-R^2 which also gets a 'variance explained' without the bias.

- L296 It may be worth mentioning the V_{Tot} as calculated from the data and V_{Tot} as calculated from the model may be quite different if the random effect structure is highly unbalanced. Which do you authors think is most relevant? The latter, presumably?

- L309 '*we offer to rely*' should read '*we often rely*'.

- L324 I think it is obvious that the character state approach will be unbiased and a mis-specified curved-parameter approach will be biased. The reason for choosing a curved-parameter approach is not lack of bias, but because it has fewer parameters and so is more precisely estimated. I think it would be more interesting to show the mean-squared error in V_{Plas} and V_{gen}. However, my guess is that the relative magnitudes of the mean-squared error will favour the character state approach when sample sizes are large, as in these simulations.

- L338 I would not use the word '*robust*' here as it has a precise statistical meaning: perhaps use '*unbiased*'

- L341 I'm not convinced R^2_{mod} should be used to distinguish the models as it doesn't have known statistical properties. If a character state approach is set up for each unique value of ϵ then a curve-parameter model is nested within it, and a straight forward likelihood ratio test could be used. Alternatively, for the fixed effects, both the curve-parameter model and the character states can be fit simultaneously (although clearly some number of the character state coefficients will be aliased) and sequential F-tests used. It would be useful to state somewhere that the character state approach can always be recovered using a polynomial with sufficient order (de Jong 1995).

- L359 Here, or elsewhere, I think it would be good to restate the important result in de Villemereuil et al. (2016) that an additive genetic model on some scale produces non-additive genetic variance on the transformed scale when the transform is non-linear. Hence, the genetic variation on the transformed scale may need partitioning into additive and non-additive variance even when the model is purely additive on the original scale. If this is not restated, I think few readers will understand why the authors are having to distinguish between broad-sense and narrow sense genetic variance.

- L411 Eq 29 I've always found this way of looking at evolutionary change in phenotype when there is plasticity a little awkward. I prefer to think of a multivariate system with the phenotype z and the reaction norm parameters as traits. Evolutionary change is then determined by the genetic covariances between z and the reaction norm parameters (which are a function of ϵ). In the example here, the selection vector would then be zero except for the element pertaining to z, but in reality there is likely to be a cost to plasticity which can be easily accommodated by having other elements of the selection vector be non-zero. This approach produces a function $\Delta\bar{z}(\epsilon)$ which needs to be averaged over the distribution of ϵ and the γ's in this paper are essentially doing this averaging, I believe. However, if there is a cost to plasticity does the approach advocated here work, and what if there is environment specific selection (the main interest from a GxE perspective) - can the two β's be replaced with β_i and β_j respectively?

- L423 problems with reference formatting.

- L474 should be '*de Jong*' not '*De Jong*'.

- L435 I would argue that if you were interested in the genetic variation in plasticity, rather than genetic variation *per se*, then the approach advocated here would fail to provide insight. It is in fact hard for me to see how a (meta) analysis of genetic variation in plasticity would not at some point have to focus on the variance in reaction norm parameters.

- L441 To use numerical integration it is assumed the environment follows a known distribution, and the parameters of that distribution are known without error. I would think conditioning on the observed environmental variables in the data, as is typically done when calculating an R^2, would be simpler and more robust to miss-specification of the environmental distribution?

References

Albecker, M. A., G. C. Trussell, and K. E. Lotterhos. 2022. A novel analytical framework to quantify co-gradient and countergradient variation. Ecology Letters 25:1521–1533.

de Jong, G. 1995. Phenotypic plasticity as a product of selection in a variable environment. American Naturalist 145:493–512.

de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey. 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. Genetics 204:1281–1294.

Johnson, P. C. 2014. Extension of nakagawa & schielzeth's R^2_{GLMM} to random slopes models. Methods in ecology and evolution 5:944–946.

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, et al. 2014. Evolutionary change in continuous reaction norms. The American Naturalist 183:453–467.