Supplementary Material

Covariance reaction norms: A flexible method for estimating complex environmental effects on trait (co)variances

Jordan S. Martin

This supplement provides more extensive details on the simulations presented in the main text, as well as additional discussion on key extensions of the basic CRN model presented in Eq. 2 of the main text.

Contents

Computational efficiency

As noted in the main text, direct prediction of large positive-definite covariance matrices is computationally challenging. Therefore, the CRN can be most efficiently estimated in Stan by using reparameterizations of the $\boldsymbol{G}_{(\boldsymbol{X})}$ and $\boldsymbol{P}_{(\boldsymbol{X})}$ matrices that are mathematically equivalent to the more intuitive parameterizations presented in Eq. 2-4. Firstly, the *p* x *p* correlation matrix $\boldsymbol{R}_{\boldsymbol{a}}$ containing all genetic (or phenotypic) correlations for *p* phenotypes can be decomposed using a Cholesky factorization such that

$$\boldsymbol{R}_{\boldsymbol{a}} = \boldsymbol{L}_{\boldsymbol{R}}\boldsymbol{L}_{\boldsymbol{R}}^{\top} \tag{S1}$$

where $\boldsymbol{L}_{\boldsymbol{R}}$ is a lower-triangular matrix with unit length rows and positive diagonal elements. These assumptions reduce the number of free parameters necessary for calculating $\boldsymbol{R}_{\boldsymbol{a}}$, as the diagonal elements of $\boldsymbol{L}_{\boldsymbol{R}}$ are determined by the off-diagonal elements of each row. Therefore, estimating $\boldsymbol{L}_{\boldsymbol{R}}$ and subsequently deriving $\boldsymbol{R}_{\boldsymbol{a}}$ using improves computational time of the model (Stan Development Team, 2023). Following previous work on the prediction of covariance matrices (Lewandowski et al., 2009; Bloome & Schrage, 2021), computational efficiency can then be further increased by decomposing $\boldsymbol{L}_{\boldsymbol{R}}$ into a vector $\boldsymbol{\omega}$ of length $\frac{p(p-1)}{2}$ containing the canonical partial correlations constitutive of all unique lower-triangular elements in this matrix. The canonical partial correlations in $\boldsymbol{\omega}$ are of the same sign as their corresponding elements in $\boldsymbol{L}_{\boldsymbol{R}}$, but their magnitudes represent residual correlations between corresponding row and column variables after regressing both on all prior occurring row variables. In the general case, the canonical partial correlation ω_u, where $u = \frac{2cp-c^2+2r-3c-2}{2}$ is the vector element corresponding to unique lower-triangular Cholesky factor $L_{R[r,c]}$ at row *r* and column *c*, is given by

$$\omega_u = \begin{cases} L_{R[r,c]}, & \text{if } c = 1 < r \\ L_{R[r,c]} \,/\, \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \tag{S2}$$

such that the Cholesky factor can in turn be derived from ω_u by

$$L_{R[r,c]} = \begin{cases} \omega_u, & \text{if } c = 1 < r \\ \omega_u * \left(1 - \sum L^2_{R[r,1:c-1]}\right)^{\frac{1}{2}}, & \text{if } 1 < c \leq r \end{cases} \quad \text{(S3)}$$

This general decomposition strategy can be adapted for the CRN model by extending each element in the vector $\boldsymbol{\omega}$ to its own vector of context-specific canonical partial correlations. Using the same approach developed in the main text (Eq. 2), environmental effects can then be specified and estimated more efficiently as predictors of the transformed canonical partial correlations

$$\begin{bmatrix} \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})_1}\right) \\ \vdots \\ \operatorname{atanh}\left(\boldsymbol{\omega}_{(\mathbf{X})\frac{p(p-1)}{2}}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_{\boldsymbol{\omega}_1} \\ \vdots \\ \boldsymbol{X\beta}_{\boldsymbol{\omega}_{\frac{p(p-1)}{2}}} \end{bmatrix} \quad \text{(S4)}$$

Applying the inverse link function $\tanh()$ and using Eq. S3 to calculate Cholesky factorized matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$, the original context-specific correlation matrices can then be derived $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ and subsequently applied to generate model predictions for estimating environmental effects on a more familiar scale. It is important to emphasize that the proposed implementation in Stan (Eq. S1-4) ensures the positive definiteness of the resulting correlation matrices $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ predicted by the CRN. Given that environmental effects are specified separately for trait correlations and variances in the CRN model (Eq. 2), the context-specific (co)variance matrices $\boldsymbol{G}_{(\boldsymbol{X})}$ derived from correlation matrices $\boldsymbol{R}_{\boldsymbol{a}(\boldsymbol{X})}$ (Eq. 1.3) will necessarily be positive definite.

Covarying environmental predictors can also reduce the efficiency and accuracy of CRN parameter estimation. To reduce the effects of collinearity, the CRN fixed effects $\boldsymbol{\beta}_{\boldsymbol{\sigma}^2}$ and $\boldsymbol{\beta}_{\boldsymbol{\omega}}$ (or $\boldsymbol{\beta}_{\boldsymbol{r}}$) can also be more efficiently estimated using a so-called thin QR factorization of the $\boldsymbol{X}$ matrix (Harville, 1997). This involves decomposing the predictor matrix $\boldsymbol{X} = \boldsymbol{Q}^*\boldsymbol{R}^*$ into an orthogonal matrix $\boldsymbol{Q}^* = \boldsymbol{Q}\sqrt{n-1}$ and upper-triangle matrix $\boldsymbol{R}^* = \frac{\boldsymbol{R}}{\sqrt{n-1}}$, estimating trait-specific regression coefficients using the orthogonal vectors $\boldsymbol{Q}^*\boldsymbol{\beta}^*$, and then returning regression coefficients appropriately scaled to the original data scale of $\boldsymbol{X}$ using $\boldsymbol{\beta} =$

$\boldsymbol{R}^{*-1}\boldsymbol{\beta}^{*}$. The QR decomposition increases efficiency by reducing posterior correlations during model sampling that would otherwise result from covariation among predictors.

Finally, the Cholesky matrices $\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X})}$ can be further used to more efficiently predict individuals' context-specific additive genetic values from the CRN model. Following previous work by (Martin & Jaeggi, 2022), this can be accomplished using a matrix normal sampling distribution (Dutilleul, 1999), which extends the vectorized multivariate normal distribution to the sampling of multivariate normally distributed matrices. Using a n x p matrix $\boldsymbol{Z}_{\boldsymbol{G}}$ of standardized individual-level additive genetic deviations (i.e. z-scores of breeding values), a lower-triangular Cholesky decomposition $\boldsymbol{L}_{\boldsymbol{A}}$ of the relatedness matrix, and a diagonal matrix $\boldsymbol{S}_{\boldsymbol{a}(\boldsymbol{X}_{\boldsymbol{n}})} = \mathrm{diag}\left(\left[\sigma_{a(X_n)_1}, \ldots, \sigma_{a(X_n)_p}\right]\right)$ of context-specific genetic standard deviations, an *n* x *p* matrix of context-specific genetic values for each phenotype can be predicted by

$$\left[\boldsymbol{a}_{(\boldsymbol{X}_{\boldsymbol{n}})_1}, \ldots, \boldsymbol{a}_{(\boldsymbol{X}_{\boldsymbol{n}})_p}\right] = \mathbf{L}_{\mathbf{A}}\boldsymbol{Z}_{\boldsymbol{G}}\left(\mathbf{S}_{\mathbf{a}(\mathbf{X}_{\mathbf{n}})}\boldsymbol{L}_{\boldsymbol{R}(\boldsymbol{X}_{\boldsymbol{n}})}\right)^{\top} \sim \mathrm{Matrix\ Normal}\left(\mathbf{0}_{\mathrm{nxp}}, \boldsymbol{A}, \boldsymbol{G}_{(\boldsymbol{X}_{\boldsymbol{n}})}\right) \quad \text{(S5)}$$

$$\rightarrow \mathrm{vec}\left(\left[\boldsymbol{a}_{(\boldsymbol{X}_{\boldsymbol{n}})_1}, \ldots, \boldsymbol{a}_{(\boldsymbol{X}_{\boldsymbol{n}})_p}\right]\right) \sim \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{G}_{(\boldsymbol{X}_{\boldsymbol{n}})} \otimes \boldsymbol{A}\right)$$

R functions are provided to facilitate these computational gains while also generating more intuitive model estimates and predictions with respect to the standard parameterizations presented in the main text. See the `CRN functions.R` file in the corresponding Github repository for further details.

Simulation-based calibration (model validation)

To provide a general validation of the proposed model, I conducted a simulation-based calibration (SBC) procedure to assess whether the CRN (**Eq. 2**) is an unbiased Bayesian estimator. SBC is a gold-standard procedure for assessing the performance of a Bayesian algorithm across a broad range of effect sizes, using synthetic datasets generated from simulated distributions of parameter values (Talts et al., 2018). This approach removes the need for arbitrarily picking a limited, discrete range of effect sizes for assessing the validity of inferences and, in so doing, reduces the risk of missing unexpected sources of bias in unexplored regions of parameter space. Instead, during SBC, generative distributions of parameter values are simulated, covering both the range of small to moderate effect sizes typically considered in standard biological applications, as well as extremely small or large values that are likely to be rare but in principle possible to observe in practice. The proposed model is then applied to the synthetic datasets generated from these simulated distributions of parameter values (priors), in turn producing distributions of estimated parameter values (posteriors). The formal correspondence between the simulated distributions of expected parameter values and the inferred distributions of estimated parameter values can then be assessed. In particular, if the proposed model implementation facilitates valid and unbiased Bayesian inference, such that the generative values are not systematically over or underestimated compared to the estimated values, the ranks of prior versus posterior values will be uniformly distributed (Talts et al., 2018). A null hypothesis of uniformity can be assessed by visualizing the difference in empirical cumulative density functions (ECDFs) with respect to the distribution of fractional ranks among the generative versus estimated values (Säilynoja, Bürkner, & Vehtari, 2022). If the difference falls within the ECDF difference 95% probability interval of the null hypothesis of uniformity, results support a valid model that generates unbiased inference across the range of generated effect sizes. See **Fig. S1a** for a visualization of the SBC procedure.

200 datasets were simulated for SBC under modest sampling conditions of 300 individuals with a single measurement for each of 3 traits. Measurements were taken in 10 environmental contexts capturing the main effects and interaction between two continuous covariates (e.g. monthly temperatures, ages, plot densities). This design was used to validate inference from the CRN in cases with minimal information available for estimation of a complex function (no repeated measures, relatively low sample size, very few environmental contexts), which is common in field applications of the animal model. Parameter values were generated using distributions appropriate for typical effect sizes in biological applications (Lemoine, 2019; McElreath, 2020), such that $\boldsymbol{\beta} \sim N(0,1)$ for RN fixed effects determining phenotypic means and genetic (co)variances, and $\boldsymbol{R_\epsilon} \sim \mathrm{LKJ}(10)$ and $\boldsymbol{\sigma_\epsilon} \sim \mathrm{exponential}(2)$ for residual correlation matrices and standard deviations respectively. Following previous work (Thomson et al., 2018), 10-generation pedigrees were simulated using variable degrees of extra-pair paternity (15-25%) and successful offspring recruitment into the breeding pool (40-70%), generating relatively sparse $\boldsymbol{A}$ matrices comparable to those typically observed in natural populations. Posterior distributions for each dataset were estimated using 2000 MCMC samples across 4 chains with 500 MCMC samples each for warmup. See the 'SBC_CRN.R' file in Github repository for full details on the simulation and model structure.

Results from the SBC analysis are visualized in **Fig. S1b**. The formal analysis using fractional ranks shows that with a 0.95+ probability, posterior inferences were not systematically upwardly or downwardly biased from the true values used to generate the data, indicating that the proposed Bayesian estimator facilitates valid inference of CRNs even under conditions of modest sampling effort and a broad range of effect sizes.

Model comparison simulation

For interpretability, the environmental variable was conceptualized as a measure of average temperature during the summer and winter months individuals were measured across 5 years, resulting in 5 * 2 = 10 environmental contexts. This relatively small number of distinct environmental contexts ensured results relevant for typical field studies lacking deep time series. Seasonal temperatures of arbitrary scale were generated for year *t* from a stationary autoregressive moving-average function of the form

$$x_t^* = \mu + 0.9\Delta x_{t-1} - 0.5\Delta x_{t-2} + 0.9\Delta\epsilon_{t-1} \quad (S6.1)$$

$$\boldsymbol{\epsilon} \sim N(0,0.1)$$

$$\mu = \begin{cases} 1, & summer \\ 0, & winter \end{cases}$$

A small cumulative value was then added to simulate weak directional change (i.e. climate warming).

$$x_t = x_t^* + 0.1t \quad (S6.2)$$

For each simulated dataset, this produced an autocorrelated time series of seasonal temperatures with a small upward directional trend (Fig. 3a), reflecting a combination of stochastic temperature fluctuations and weak climate warming across years. The temperature variable $\boldsymbol{x}$ was subsequently standardized to unit variance to charitably compare the performance of the CRN and character state models with binned environmental states (summer = $x_t > 0$, winter = $x_t < 0$), given that $\bar{x}_{\text{summer}} - \bar{x}_{\text{winter}} = 1$. This allowed for more direct comparison of the expected change in genetic (co)variance under a 1-unit change between the continuous temperature variable and binned seasonal state.

Genetic covariances were simulated from a CRN model for two Gaussian phenotypes such that

$$\begin{bmatrix} \boldsymbol{z_1} \\ \boldsymbol{z_2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} + \boldsymbol{\alpha}_{(X)_1} + \boldsymbol{\epsilon_1} \\ \mathbf{0} + \boldsymbol{\alpha}_{(X)_2} + \boldsymbol{\epsilon_2} \end{bmatrix} \quad (S7)$$

147 $$\begin{bmatrix}\boldsymbol{a}_{(X)_1}\\ \boldsymbol{a}_{(X)_2}\end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(X)} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(X_n)}: \begin{bmatrix}\sigma^2_{a(X_n)_1} & r_{a(X_n)_{1,p}}\sigma_{a(X_n)_1}\sigma_{a(X_n)_p}\\ & \sigma^2_{a(X_n)_2}\end{bmatrix}$$

148 $$\text{small effect size: } \begin{bmatrix}\log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(X)_1}\right)\\ \log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(X)_2}\right)\end{bmatrix} = \begin{bmatrix}[\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.3\end{bmatrix}\\ [\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.3\end{bmatrix}\end{bmatrix};\quad \begin{bmatrix}\operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(X)_{1,2}}\right)\\ \operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(X)_{p-1,p}}\right)\end{bmatrix} = \begin{bmatrix}[\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}0\\ 0.1\end{bmatrix}\\ [\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}0\\ 0.1\end{bmatrix}\end{bmatrix}$$

149 $$\text{medium effect size: } \begin{bmatrix}\log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(X)_1}\right)\\ \log\left(\boldsymbol{\sigma}^2_{\boldsymbol{a}(X)_2}\right)\end{bmatrix} = \begin{bmatrix}[\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.6\end{bmatrix}\\ [\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}-1.2\\ 0.6\end{bmatrix}\end{bmatrix};\quad \begin{bmatrix}\operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(X)_{1,2}}\right)\\ \operatorname{atanh}\left(\boldsymbol{r}_{\boldsymbol{a}(X)_{p-1,p}}\right)\end{bmatrix} = \begin{bmatrix}[\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}0\\ 0.2\end{bmatrix}\\ [\boldsymbol{1} & \boldsymbol{x}]\begin{bmatrix}0\\ 0.2\end{bmatrix}\end{bmatrix}$$

150 $$\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma});\ \boldsymbol{\Sigma} = \boldsymbol{SRS}$$

151 $$\boldsymbol{R} \sim LKJ(10)$$

152 $$\mathbf{S} = \operatorname{diag}\left(\sqrt{0.7}, \sqrt{0.7}\right)$$

153 Mean trait changes with temperature were ignored for simplicity given the purposes of the simulation,
154 though this is of course biologically unrealistic. Note that $\exp(-1.2) = \boldsymbol{\sigma}^2_{\boldsymbol{a}(\boldsymbol{x}=\boldsymbol{0})} \approx 0.3$ and $\tanh(0) =$
155 $\boldsymbol{r}_{\boldsymbol{a}(\boldsymbol{x}=\boldsymbol{0})} = 0$, so that for the average temperature (x = 0), traits exhibited modest heritability $h^2 = \frac{0.3}{0.3+0.7}$
156 and were uncorrelated. Cross-environment correlations were fixed to r = 0.8 by simulating 10 correlated
157 standardized genetic values across i individuals for each of the 2 traits, constructing 10 $\boldsymbol{Z}_{\boldsymbol{G}}$ matrices (i x 2)
158 from these correlated standard normal values, and subsequently scaling them using context-specific $\boldsymbol{G}_{(\boldsymbol{X})}$
159 matrices following **Eq. S5**. Relatedness matrices $\boldsymbol{A}$ were simulated following the same procedure used for
160 the SBC as described above. See the 'methods comparison.R' script in the accompanying Github repository
161 for further details.

162 It may be argued that this simulation unfairly privileges the CRN model over the random
163 regression, given that the pattern of genetic change is not simulated directly from an individual-level model
164 with corresponding intercepts and slopes. However, there are two key points to consider. Firstly, I have only
165 assessed performance with respect to positive temperature change, i.e. $\Delta G_{1,2} = 0.04 = G_{1,2(x=1)} -$

$G_{1,2(x=0)}$ (Fig. 3a), thus effectively ignoring the larger bias expected for negative values, where the random regression model will incorrectly infer symmetric change due to predicting quadratic change in (co)variances. Secondly, given the effect sizes simulated, linear slopes can effectively approximate the pattern of individual trait change that would be expected, over the simulated range of the temperature variable, to generate the observed pattern of genetic change simulated from the CRN. In particular, focusing on a single trait z for the small effect size condition and ignoring non-genetic effects, assume the data were simulated by a random regression model of the form

$$z_{1i} = 1 + \boldsymbol{b}e^{0.15x} \quad (\mathbf{S8.1})$$

$$\boldsymbol{b} \sim N(0,1\boldsymbol{A})$$

where $\boldsymbol{b}$ are additive genetic random slopes. The expected variance of the trait at a given value of *x* is

$$\sigma^2_{z_1} = \sigma^2_{\boldsymbol{b}}\exp(0.3\mathrm{x}) = \exp(0.3x) \quad (\mathbf{S8.1})$$

Consistent with the simulated variance CRN (Eq. S7). Plotting this function for different values of $\boldsymbol{b}$ over the typical range of standardized temperature values $[-1,1]$, as shown in Figure S2 below**,** demonstrates that this function can be well approximated by a standard random regression model with genetically varying intercepts and linear slopes. Of course, when two models make distinct empirical assumptions about the functional form of trait change, it is difficult to use synthetic data to make impartial comparisons between them. A valuable target for future empirical research will, therefore, be to more directly quantify the curvature of GxE for trait (co)variances across many traits, environments, and clades. This will provide general heuristics for the conditions under which specific functional forms are likely to provide better approximations to the underlying reality, which can then be incorporated into specific CRN models.

Estimating cross-environment correlations

Cross-environment correlations are determined by the genetic variance in and covariance among the intercepts and slopes of individuals' mean reaction norms (Brommer, 2013; Mitchell & Houslay, 2021), providing crucial information for predicting evolutionary change when organisms exhibit heritable variation in their plasticity toward the environment (Martin et al., 2024). The basic CRN model (**Eq. 2**) does not directly estimate cross-environment correlations. However, it can be readily extended to do so with an appropriate breeding design or when repeated measures are available on the same individuals across environments.

Note that when repeated measures are available, such that the number of subjects $i < n$, the matrix of standardized individual-level additive genetic deviations $\boldsymbol{Z_G}$ can instead be modeled as an $i\ x\ p$ matrix that is constant across environmental contexts. This implies constant rank-order among subjects (cross-environmental genetic correlations $r = 1$) and thus the absence of individual-level variation in plasticity (Brommer, 2013) with respect to the traits being measured. This approach will generally be most applicable when individuals are measured repeatedly within rather than across environmental contexts, such that only population-level GxE can be estimated. Alternatively, when individuals are measured repeatedly across rather than within environmental contexts, the most flexible and straightforward approach to allow for $r < 1$ is to keep $\boldsymbol{Z_G}$ as an $n\ x\ p$ matrix just as in the CRN model without repeated measurements. This freely estimates subjects' standardized values and rank-order in each unique context. Cross-environment correlations can then be manually computed from the posterior distributions of these context-specific values, allowing for flexible estimation of any within- or among-individual correlations of interest (e.g. genetic assortment coefficients, Martin & Jaeggi, 2022), even in cases where only a subset of individuals are repeated measured across contexts. Cross-environment correlations can also be directly estimated in the model by expanding $\boldsymbol{Z_G}$ to an $n\ x\ p * c$ matrix for c contexts, as in a standard character state model, while still allowing for the CRN to predict genetic (co)variances among traits.

Genetic prediction

If one is interested in predicting genetic values for all individuals across all contexts, a more general form is to specify $\boldsymbol{Z_G}$ as a C-dimensional array of $i \; x \; p$ matrices, estimating individuals' expected values for all contexts, even those they weren't observed in. In other words, one can simply extend the basic model to include standardized (unscaled) random effects for all individuals in all possible contexts. This can be useful for various purposes, including estimating cross-environment correlations when some subjects are only observed in a subset of environmental contexts. However, this strategy can become cumbersome very rapidly for models with many environmental contexts and will result in much slower sampling. This motivates use of random regression techniques for high-dimensional and continuously varying environments.

Random regression CRN

Random regression and CRN models can be synthesized to determine how the cross-environment correlations among individuals' genetic values as well as the genetic (co)variances among traits are shaped by nested patterns of environmental variation. Random individual-level slopes can be introduced to the CRN model so that the CRN describes changes in the (co)variances of the intercepts and slopes governing RNs of trait means. For instance, empiricists may be interested in testing theoretical predictions of how the genetic integration between individuals' mean trait value and plasticity to the environment changes across developmental or social contexts (Kraft et al., 2006; Stamps et al., 2018; Dingemanse et al., 2020; Bucklaew & Dochtermann, 2021; Martin et al., 2023). A random regression CRN can be implemented under a proper experimental design for detecting GxE and/or with repeated measurements, where individuals' breeding values for environmental slopes can be estimated from observations of related individuals' trait values across at least two or more environmental contexts. To do so, new vectors and matrices need to be introduced: *v*i x 1* vectors $\boldsymbol{u}$ for each phenotype containing *v* random effects (intercepts and slopes) for *i* individuals, and *n* x *v*i* block diagonal design matrices $\boldsymbol{W}$ indexing repeated measurements and scaling the

v random effects for *i* individuals across *n* total measurements of each phenotype. The random regression CRN is then given by

$$\begin{bmatrix} g_{z_1}(\boldsymbol{z_1}) \\ \vdots \\ g_{z_p}(\boldsymbol{z_p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta}_1 + \boldsymbol{Wu}_{(\boldsymbol{X})_1} + \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{X\beta}_p + \boldsymbol{Wu}_{(\boldsymbol{X})_p} + \boldsymbol{\epsilon_p} \end{bmatrix} \quad (S9)$$

$$\begin{bmatrix} \boldsymbol{u}_{(\boldsymbol{X})_1} \\ \vdots \\ \boldsymbol{u}_{(\boldsymbol{X})_p} \end{bmatrix} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G}_{(\boldsymbol{X})} \otimes \boldsymbol{A}\right);\ \boldsymbol{G}_{(\boldsymbol{X_n})}: \begin{bmatrix} \sigma^2_{\alpha(X_n)_1} & \cdots & r_{a_{X_{I1}}(X_n)_1, b_{X_{Ib}}(X_n)_p} \sigma_{a_{X_{I1}}(X_n)_1} \sigma_{b_{X_b}(X_n)_p} \\ & \ddots & \vdots \\ & & \sigma^2_{\beta_b(X_n)_p} \end{bmatrix}$$

Note that the design matrix $\boldsymbol{W} = \text{blockdiag}(\boldsymbol{X_{v1}}, \ldots, \boldsymbol{X_{vi}})$ is a block diagonal matrix containing repeated observations of individuals 1 to *i* from the subset of *v* columns in the full environmental matrix $\boldsymbol{X}$ over which individual intercepts $\boldsymbol{a}_{(\boldsymbol{X})_p}$ and slopes $\boldsymbol{\beta}_{\boldsymbol{1}(\boldsymbol{X})_p}, \ldots, \boldsymbol{\beta}_{\boldsymbol{b}(\boldsymbol{X})_p}$ are defined in the model for trait *p*. The process of prediction for the elements in $\boldsymbol{G}_{(\boldsymbol{X})}$ is equivalent to Eq. 2, though the total number of parameters to estimate in a full random regression CRN model expands to $k = b\frac{vp(vp+1)}{2}$, where b is the number of environmental CRN parameters and v is the number of individual effects. A phenotypic version of the random regression CRN can also be implemented following Eq. 3.

By allowing genetic variation in response to the environment to also change across environments, the random regression CRN can simultaneously quantify both population- and individual-level parameters shaping environmental effects on the $\boldsymbol{G}$ matrix, providing more accurate predictions about the dual consequences of developmental and contextual plasticity across multiple scales. Pragmatically, this model will be particularly useful when individuals are only measured across a subset of relevant environmental contexts. For instance, subjects may be repeatedly measured under varying microclimatic and resource conditions within their local patch, providing information to estimate the average within-patch (co)variation of intercepts and slopes, without experiencing among-patch variation in community composition and habitat quality, which may further shape genetic (co)variation of these parameters and thus the magnitude of GxE across environments.

Link functions for trait variances

Using the $\log(\sigma^2)$ function for trait variances facilitates unbiased inference on the latent scale of the CRN model (Fig. S1), as well as the observed scale when it is characterized by exponential change (over the range of prediction) in response to the environment (Fig. 3). The log link is also a commonly used and familiar function that facilitates intuitive interpretation. However, this function may not always be the best choice when estimating a CRN. Below I consider two alternatives, one which is generally not recommended (square root) and another which is likely to be much more broadly applicable (inverse softplus).

Square root link

The Gaussian linear random regression model assumes that trait variances change exponentially with respect to the environment, which in turn implies linear change in genetic standard deviations for positive values (with negative values leading in the extreme to predictions of improper negative standard deviations; see Fig. S3 for a comparison). As discussed in the main text, this is unlikely to be a realistic assumption for many traits, given its implication of symmetric change in genetic variance across positive and negative values of the environment. However, when quadratic change is the true functional form for the variance, one can instead use a square root link function directly on the genetic variance, such that $\sqrt{\boldsymbol{\sigma^2}} = |\boldsymbol{X}|\boldsymbol{\beta_\sigma}$ where predictors are constrained to be positive $\boldsymbol{\beta_\sigma} > \boldsymbol{0}$ in the model likelihood to ensure identifiability. This implies that $(|\boldsymbol{X}|\boldsymbol{\beta_\sigma})^2 = \sigma^2$, so that the variance changes quadratically as a function of the positive linear environmental effects on the standard deviation. The key issue here is ensuring that $\boldsymbol{X}\boldsymbol{\beta_\sigma} > 0$, so this must be accomplished by scaling both the linear predictors and the environmental values (by taking the absolute value) to be strictly positive, unless one is modeling data where there is no risk of predicting impossible negative values. To the degree that one is certain this model is appropriate, it will generally be much more straightforward and interpretable to fit a standard or CRN random regression model (Eq. S9), and thus I cannot recommend it for general use outside of very specific cases (e.g. when repeated measures are unavailable but prior work informs a strong expectation of quadratic change).

Inverse softplus link

Another less commonly employed link function may be more useful for modelling genetic variance in the CRN model. An important limitation of the log link is that, for larger effect sizes, it can be prone to upwardly bias estimates on the original data scale, due to the asymmetric influence of estimation error. This follows from Jensen's inequality, where for some convex function f and random variable x

$$f\big(\mathrm{E}(x)\big) \leq \mathrm{E}\big(f(x)\big) \tag{S10.1}$$

which implies that

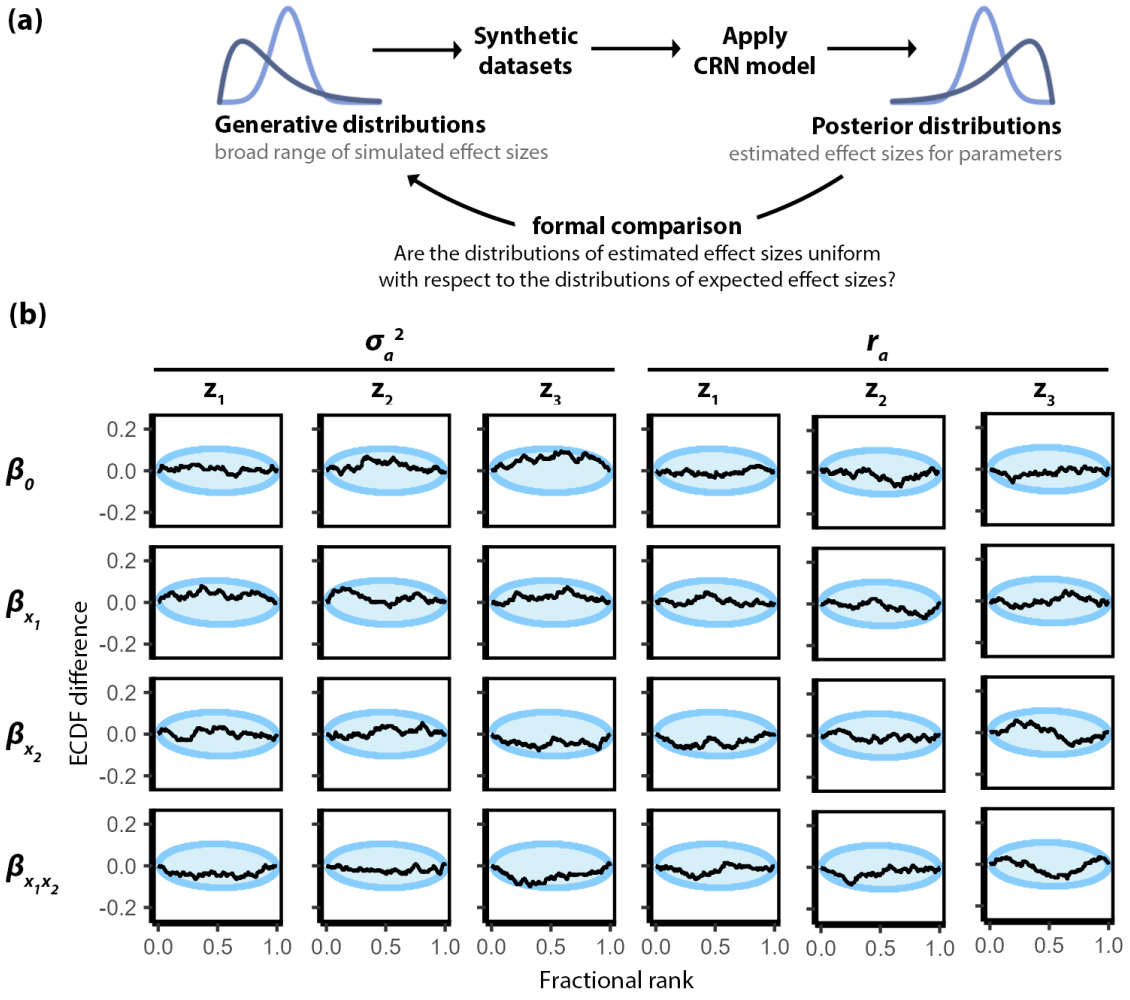
$$f(\sigma^2) < \mathrm{E}\left(\exp\big(\widehat{\sigma^2}\big)\right) \tag{S10.2}$$

$$\widehat{\sigma^2} \sim N(\sigma^2, \delta)$$

for the true trait variance σ^2 estimated by $\widehat{\sigma^2}$ with random Gaussian error of magnitude δ. In other words, even if error is truly random and thus the expected estimate is unbiased on the log scale, the expected variance estimated on the original scale $\exp\big(\widehat{\sigma^2}\big)$ will still tend to be upwardly biased from the true value, due to application of the convex exponential inverse link function. The importance of this upward bias will be contingent on factors such as the sample size, which will generally decrease error in estimates, as well as the degree to which one is focused on unbiasedly estimating CRN model coefficients on the link scale or predicting exact magnitudes of change in genetic variance on the original data scale. The former will generally be of greater importance for basic research, where theory is most often tested by qualitative predictions (e.g. covariances will become more positive/negative) rather than exact quantitative predictions. The log link should, therefore, be a fine choice for most purposes. One can always hedge their bets by expecting that there may be a small upward bias in original scale predictions in the presence of high statistical uncertainty, while being confident that latent scale predictions and thus the quantities most often used for hypothesis testing (e.g. CRN parameters) are expected to be unbiasedly estimated.

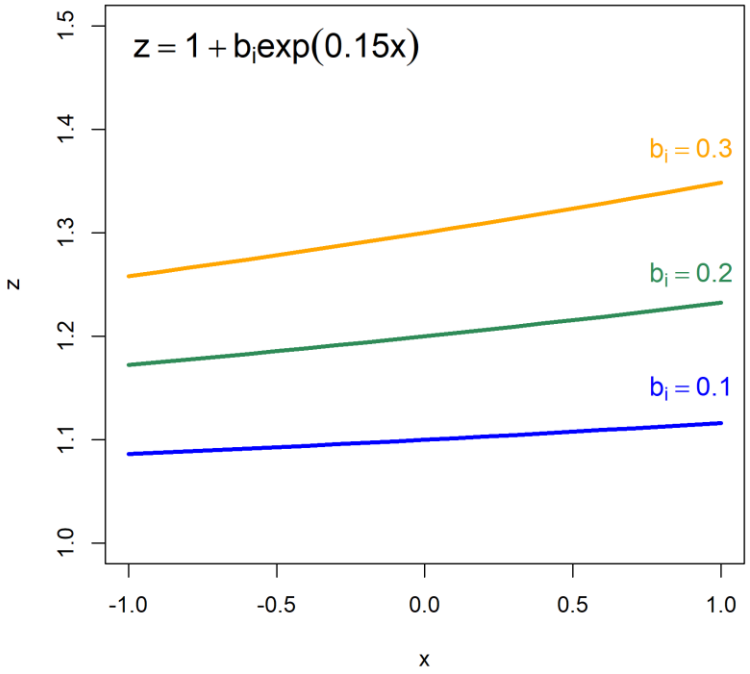
In some cases, however, it may be desirable to use the inverse softplus link function, where $\log\left(exp\left(\boldsymbol{\sigma^2}\right)-1\right)=\boldsymbol{X\beta_{\sigma^2}}$ predicts the latent scale values and the softplus function $\boldsymbol{\sigma^2}=\log\left(1+exp\left(\boldsymbol{X\beta_{\sigma^2}}\right)\right)$ returns the variance on the original scale. Latent predictions from the softplus tend to scale much less convexly with respect to the original data scale (**Fig. S3**). The clear benefit of this link function is, therefore, that stochastic estimation error on the link scale tends to generate less upward bias with respect to the original variance and standard deviation scales, in comparison to the log link. Moreover, because the softplus approximates the identity function as it moves further away from 0 variance, it provides a natural solution for flexibly testing distinct functional forms for the variance components. For instance, one can specify a 2nd-order polynomial on the link scale to more directly test for approximately quadratic change. Note that the softplus can be implemented in Stan using the 'log1p_exp' function.

Figure S1. Simulation-based calibration of the CRN model.



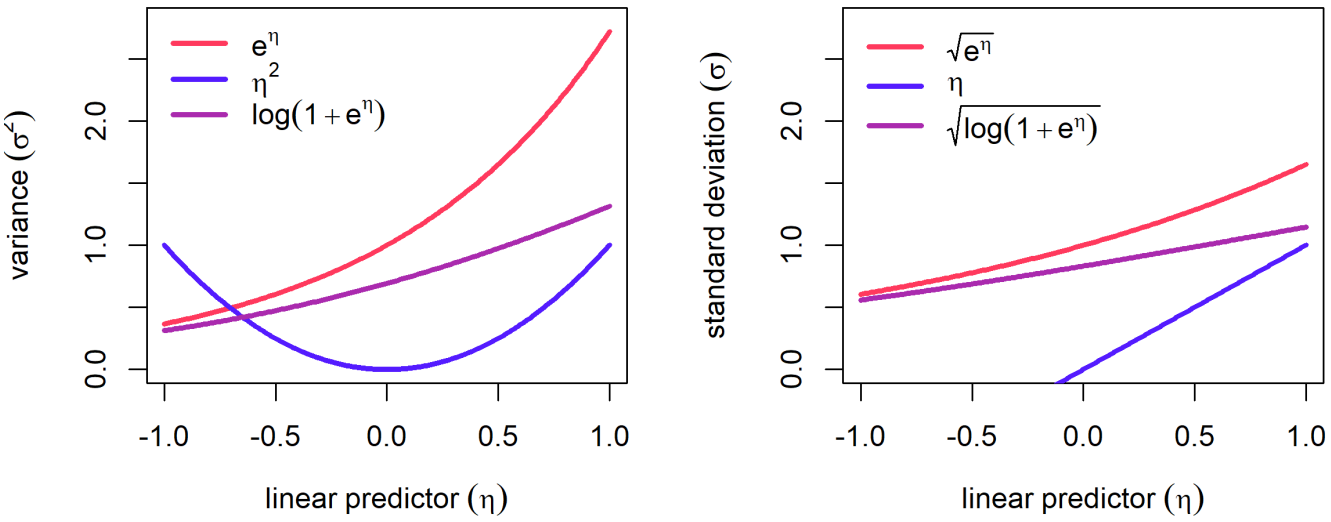
Footnote. Results are shown for SBC analysis of 200 simulated datasets of 3 traits under minimal sampling conditions (N = 300 / 10 environmental contexts) generated from prior distributions defined over the parameters of the quantitative genetic CRN model (Eq. 2). (a) Conceptual overview of the SBC procedure. (b) The CRN contained four parameters for each genetic variance (σ_α^2) and correlation (r_α): β_0 for the trait-specific intercepts, β_{x_1} and β_{x_2} for the main effects of two continuous and independently distributed environments, and $\beta_{x_1x_2}$ for the interaction effect of these continuous environments. Plots show the difference (y-axis) between the empirical cumulative density functions (ECDFs) for CRN parameters from the generative prior distributions used to simulate datasets and the ECDFs of the estimated posterior distributions across datasets. This difference is shown by the black line and plotted as a function of the relative fractional rank (x-axis) of the simulated values in comparison to inferred values. Blue ellipses show regions providing 0.95+ probability of uniformity between the ECDFs of the simulated and estimated parameter distributions, providing support for a well-calibrated model without systematic bias (Talts et al., 2018). Therefore, while stochastic fluctuations are expected at computationally efficient sample sizes, black lines should remain within the blue ellipses across fractional ranks if the model generates unbiased posterior estimates of parameter values, with respect to the prior simulated values. Consistent deviations of the black line beyond the blue ellipse provide statistical evidence of bias in the region of parameter space indicated by the fractional ranks. For instance, if a model systematically underestimates parameter values, we expect the black lines to peak outside the blue ellipses at high fractional ranks, indicating that prior values were systematically larger than inferred estimates.

Figure S2. Individual reaction norms generating exponential change.



Footnote. Examples of individual-level responses (Eq. S6) that generate the exponential change in genetic variance assumed by the CRN model used to simulate datasets for the model comparison (Eq. S7). The true function (black text) incorporates Gaussian random slopes $\boldsymbol{b}$ determining how individuals' trait expression (z) changes in response to temperature (x). Three lines of varying color are shown for individuals (i) with differing random slope values. The visualization demonstrates that over the typical range of temperature values used in the simulation, individual reaction norms from this exponential function could be well approximated by a standard linear random regression model.

Figure S3. Comparison of link functions for the genetic variance.



Footnote. The relationship is plotted between the value of the linear predictor $\eta = \boldsymbol{X_n\beta}$ (x-axis) and the original scale variance (left plot) and standard deviation (right plot) as a function of assuming a log link (exponential for original scale; red), square root link (quadratic change; blue), or inverse softplus link (purple). As can be seen, the inverse softplus link exhibits much less convexity than the log link on the original scale, facilitating approximately sublinear prediction of the variance and standard deviation for small values (the function becomes increasingly linear with larger values).

Table S1. Summary of meerkat CRN parameter posterior distributions.

Regression coefficient	variance reaction norm $\boldsymbol{\beta}_{\sigma^2_\alpha}$		correlation reaction norm $\boldsymbol{\beta}_{r_\alpha}$	
	median	$p_{+/-}$	median	$p_{+/-}$
foraging and feeding pups (FD)			FD ~ BS	
age	0.19	0.81	-0.34	0.98
sex	0.10	0.62	0.31	0.90
dominance status	1.10	1.00	0.17	0.77
age * sex	-0.07	0.61	0.11	0.70
age * dominance	-0.10	0.64	0.25	0.79
sex * dominance	-0.36	0.84	-0.28	0.80
age * sex * dominance	-0.17	0.65	-0.65	0.93
group size	0.20	1.00	-0.12	0.98
group size^2	0.21	1.00	-0.04	0.73
babysitting (BS)			FD ~ GD	
age	0.96	1.00	-0.21	0.90
sex	-0.21	0.75	0.15	0.77
dominance status	-0.02	0.52	0.19	0.80
age * sex	-0.13	0.66	-0.01	0.52
age * dominance	-0.85	0.99	0.34	0.92
sex * dominance	0.76	0.96	-0.20	0.77
age * sex * dominance	-0.01	0.56	-0.10	0.60
group size	-0.12	0.97	-0.10	0.98
group size^2	0.08	0.87	0.11	0.99
vigilant guarding (GD)			BS ~ GD	
age	-0.19	0.94	-0.16	0.77
sex	0.12	0.77	0.32	0.96
dominance status	0.49	0.99	0.23	0.85
age * sex	-0.12	0.81	0.30	0.96
age * dominance	0.47	0.99	0.15	0.71
sex * dominance	-0.01	0.52	-0.37	0.84
age * sex * dominance	0.65	0.98	-0.13	0.60
group size	0.02	0.72	0.07	0.88
group size^2	0.05	0.94	0.05	0.79

Footnote. Posterior distributions of CRN parameters (regression coefficients) for the genetic variances ($\boldsymbol{\beta}_{\sigma^2_\alpha}$) and genetic correlations ($\boldsymbol{\beta}_{r_\alpha}$) among three meerkat social behaviors: foraging and pup feeding (FD), babysitting (BS), and vigilant guarding (GD). Posteriors are summarized by their median and the probability of a directional effect ($p_{+/-}$). Note that $p_{+/-}$closer to 1 provide stronger support for a positive or negative effect, contingent on the sign of the median effect size. Reference categories for sex and dominance are female and subordinate.

References

Bloome, D., & Schrage, D. (2021). Covariance Regression Models for Studying Treatment Effect Heterogeneity Across One or More Outcomes: Understanding How Treatments Shape Inequality. *Sociological Methods & Research*, *50*(3), 1034–1072.

Brommer, J., Class, B., & Covarrubias-Pazaran, G. (2019). Multivariate Mixed Models in Ecology and Evolutionary biology: Inferences and implementation in R. In *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/hs38a

Bucklaew, A., & Dochtermann, N. A. (2021). The effects of exposure to predators on personality and plasticity. *Ethology*, *127*(2), 158–165.

Dingemanse, Niels J., Barber, I., & Dochtermann, N. A. (2020). Non-consumptive effects of predation: does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? *Ecology Letters*, *23*(1), 107–118.

Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, *64*(2), 105–123.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag.

Kraft, P. G., Wilson, R. S., Franklin, C. E., & Blows, M. W. (2006). Substantial changes in the genetic basis of tadpole morphology of Rana lessonae in the presence of predators. *Journal of Evolutionary Biology*, *19*(6), 1813–1818.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Martin, J. S., & Jaeggi, A. V. (2022). Social animal models for quantifying plasticity, assortment, and selection on interacting phenotypes. *Journal of Evolutionary Biology*, *35*(4), 520–538.

Martin, Jordan S., Jaeggi, A. V., & Koski, S. E. (2023). The social evolution of individual differences: Future directions for a comparative science of personality in social behavior. *Neuroscience and Biobehavioral Reviews*, *144*, 104980.

Martin, J. S., Westneat, D., Wilson, A. J., Dingemanse, N. J., & Araya-Ajoy, Y. (2024). Frequency-dependence favors social plasticity and facilitates socio-eco-evolutionary feedback in fluctuating environments. *EcoEvoRxiv*. https://doi.org/10.32942/X2KW5R

McElreath, R. (2020). *Statistical Rethinking; A Bayesian Course with Examples in R and Stan; Second Edition*. CRC Press.

Mitchell, D. J., & Houslay, T. M. (2021). Context-dependent trait covariances: how plasticity shapes behavioral syndromes. *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology*, *32*(1), 25–29.

Säilynoja, T., Bürkner, P.-C., & Vehtari, A. (2022). Graphical test for discrete uniformity and its applications in goodness-of-fit evaluation and multiple sample comparison. *Statistics and Computing*, *32*(2). https://doi.org/10.1007/s11222-022-10090-6

Stamps, J. A., Biro, P. A., Mitchell, D. J., & Saltz, J. B. (2018). Bayesian updating during development predicts genotypic differences in plasticity. *Evolution; International Journal of Organic Evolution*, *72*(10), 2167–2180.

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual 2.33*.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. In *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/1804.06788

Thomson, C. E., Winney, I. S., Salles, O. C., & Pujol, B. (2018). A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PloS One*, *13*(10), e0197720.