Consider P traits $(Y_1, \ldots, Y_P)$, where number of observations of each trait $Y_i = N$. When stacked, Y is a $NP \times 1$ vector. We won't consider covariates other than sites since we assume they're orthogonal and can just project them away. We consider the site covariates $X_{n \times \#sites}$ and site effects for each phenotype γ_j which are each $P \times 1$ vectors. Lastly the error ($\epsilon \sim N(0, \sigma^2)$ is scaled for each site-phenotype combination, where each δ_j is an $\#nsites$ vector that scales the error variancee

1 The Model

$$
\begin{aligned}
Y &= (Y_1, \ldots, Y_P) \\
Y &= X_g\beta + X_s\gamma + \epsilon \\
vec(Y) &= (\beta_g' \otimes I_n)vec(X_G) + (\gamma_s' \otimes I_n)vec(X_s) + vec(\epsilon) \\
var(vec(\epsilon)) &= \Sigma_e \otimes I_n
\end{aligned}
$$

Both methods will be GRM based so

$$
var(vec(X_g\beta)) = \Sigma_g \otimes A = \begin{bmatrix} \sigma^2_{g1}A & \ldots & \rho_{g,1P}A \\ \vdots & \ddots & \vdots \\ \rho_{g,P1}A & \ldots & \sigma^2_{gP}A \end{bmatrix}
$$

$$
Var(Y) = \begin{bmatrix} Var(Y_1) & \ldots & \Sigma_{1P} \\ \vdots & \ddots & \vdots \\ \Sigma_{P1} & \ldots & Var(Y_P) \end{bmatrix}
$$

Considering just two phenotypes we have

$$
\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N(\begin{bmatrix} X_s\gamma_1 \\ X_s\gamma_2 \end{bmatrix}, \begin{bmatrix} \sigma^2_{g_1}A + \sigma^2_{e_1}I & \rho_gA + \rho_eI \\ \rho_gA + \rho_eI & \sigma^2_{g_2}A + \sigma^2_{e_2}I \end{bmatrix})
$$

2 AdjHE RE method

Treating site effects as random, and then no covariance between site, genetic, or error we get

$$
\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N(0, \begin{bmatrix} \sigma^2_{g_1}A + \sigma^2_{s_1}S + \sigma^2_{e_1}I & \rho_gA + \rho_sS + \rho_eI \\ \rho_gA + \rho_sS + \rho_eI & \sigma^2_{g_2}A + \sigma^2_{s_2}S + \sigma^2_{e_2}I \end{bmatrix})
$$

This means

$$
Cov(Y_1, Y_2) = \rho_gA + \rho_sS + \rho_eI
$$

So the covariance between different traits has parts based in the similarities in the genetics, site, and expeirmental error.

3 ComBat

Long story short, CovBat technique only addresses the covariance due to the sites.

First they use Combat to residualize by subtracting the emprical Bayes estimators of the site mean (γ_S^*) and variance (δ_s^*). And we know that they emprical bayes estimators are consistent ($\gamma_S^* \xrightarrow{p} \gamma_S, \delta_S^* \xrightarrow{p} \delta_S$). Looking at the combat adjustment

$$Y^{combat} = (Y - X_s\gamma^*)\delta^{*-1}$$

We have

$$X_s\beta_s^* \xrightarrow{p} X_s\beta_s$$

Therefore a projection defined by the site means we have

$$Q_s^* \xrightarrow{p} Q_s = X_s(X_s'X_s)^{-1}X_s'$$

However, since $X \not\perp PC$ we run into the problem that

$$Q_sY = Q_sX_g\beta + \epsilon$$

Assuming that we have a perfectly balanced stduy s.t. $PC \perp X$ Then we'd have

$$Q_sY = X_g\beta + \epsilon \sim N(0, \sigma_g^2A + \sigma_e^2)$$

$$\delta^* \xrightarrow{p} \delta \overset{CMT}{\therefore} \delta^{*-1} = diag(1/\delta_j^*) \xrightarrow{p} \delta^{-1}$$

By Slutsky's theorem and CMT and assuming no differences ($\delta = I$)

$$Y^{combat} \xrightarrow{d} \epsilon \sim N(0, \sigma_g^2A + \sigma_e^2I)$$

4 Covbat

Taking the residuals from Combat Y^{combat} (which have mean 0) we denote the covariance matrices of each residualized phenotype as $var(Y_j^{combat}) = \Sigma_j$. They

take a PCA decomposition of the residualized phenotypes to get the covariance matrix

$$\Sigma = \sum_{k=1}^{p} \lambda_k \phi_k \phi_k^T$$

the residualzied phenotypes are expressed using the coordinates (η) along each of the first p eigenvectors. However, because part of the covariance contains the GRM, this would affect the heritability estimate.