agronomy

Supplemental Material

Supplementary Table S1: Marginal genetic accuracy and standard error by delta ratios and treatments.

	9:1	3:1	1:1	1:3	1:9	AVG
I	0.647 (0.002)	0.630 (0.003)	0.600 (0.009)	0.531 (0.020)	0.438 (0.029)	0.569 (0.015)
A	0.783 (0.001)	0.779 (0.001)	0.760 (0.005)	0.716 (0.012)	0.660 (0.020)	0.739 (0.009)
G	0.823 (0.001)	0.821 (0.002)	0.803 (0.006)	0.760 (0.013)	0.703 (0.022)	0.782 (0.009)
NS	0.751 (0.053)	0.737 (0.058)	0.695 (0.065)	0.612 (0.079)	0.509 (0.090)	0.661 (0.036)
RC	0.750 (0.054)	0.736 (0.059)	0.695 (0.067)	0.612 (0.082)	0.512 (0.093)	0.661 (0.036)
M	0.753 (0.052)	0.748 (0.055)	0.730 (0.059)	0.683 (0.068)	0.610 (0.081)	0.705 (0.028)
NG	0.749 (0.053)	0.748 (0.056)	0.732 (0.059)	0.689 (0.069)	0.624 (0.081)	0.708 (0.027)
SD	0.752 (0.053)	0.749 (0.056)	0.734 (0.059)	0.701 (0.064)	0.650 (0.075)	0.717 (0.025)
AR	0.750 (0.054)	0.741 (0.060)	0.727 (0.062)	0.691 (0.066)	0.646 (0.077)	0.711 (0.026)
GK	0.751 (0.055)	0.744 (0.061)	0.733 (0.061)	0.695 (0.065)	0.648 (0.078)	0.714 (0.027)

Supplementary Table S2: Marginal genetic accuracy and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
NS	0.525 (0.059)	0.709 (0.037)	0.749 (0.039)
RC	0.521 (0.059)	0.711 (0.037)	0.752 (0.037)
MA	0.581 (0.037)	0.746 (0.022)	0.788 (0.021)
NG	0.584 (0.034)	0.749 (0.019)	0.792 (0.018)
SD	0.598 (0.027)	0.756 (0.016)	0.799 (0.015)
AR	0.585 (0.027)	0.753 (0.015)	0.794 (0.015)
GK	0.589 (0.027)	0.752 (0.017)	0.801 (0.014)

Supplementary Table S3: Marginal spatial accuracy and standard error by delta ratios and treatments.

	9:1	3:1	1:1	1:3	1:9	AVG
I	0.422 (0.056)	0.569 (0.064)	0.704 (0.061)	0.787 (0.067)	0.830 (0.072)	0.662 (0.039)
A	0.444 (0.058)	0.606 (0.058)	0.727 (0.063)	0.799 (0.068)	0.837 (0.073)	0.683 (0.038)
G	0.457 (0.058)	0.616 (0.059)	0.741 (0.064)	0.804 (0.068)	0.840 (0.073)	0.692 (0.037)
RC	0.241 (0.009)	0.342 (0.009)	0.423 (0.007)	0.471 (0.004)	0.490 (0.002)	0.394 (0.025)
MA	0.457 (0.002)	0.612 (0.001)	0.717 (0.001)	0.776 (0.001)	0.800 (0.000)	0.672 (0.034)
NG	0.559 (0.014)	0.698 (0.011)	0.798 (0.008)	0.865 (0.006)	0.905 (0.003)	0.765 (0.034)
SD	0.535 (0.014)	0.702 (0.011)	0.819 (0.008)	0.896 (0.005)	0.940 (0.004)	0.778 (0.039)
AR	0.293 (0.012)	0.516 (0.045)	0.765 (0.031)	0.878 (0.009)	0.939 (0.004)	0.678 (0.065)
GK	0.561 (0.013)	0.713 (0.011)	0.823 (0.010)	0.896 (0.006)	0.938 (0.003)	0.786 (0.037)

Supplementary Table S4: Marginal spatial and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
NS	0.382 (0.048)	0.396 (0.046)	0.403 (0.044)
RC	0.672 (0.063)	0.672 (0.063)	0.672 (0.063)
MA	0.749 (0.066)	0.767 (0.062)	0.779 (0.059)
NG	0.762 (0.077)	0.784 (0.072)	0.790 (0.071)
SD	0.640 (0.127)	0.689 (0.119)	0.706 (0.118)
AR	0.770 (0.071)	0.788 (0.067)	0.800 (0.066)

Supplementary Table S5: Marginal selection accuracy (top 10%) and standard error by delta ratios and treatments.

	9:1	3:1	1:1	1:3	1:9	AVG
I	0.456 (0.002)	0.437 (0.002)	0.417 (0.006)	0.361 (0.014)	0.306 (0.018)	0.395 (0.010)
A	0.563 (0.001)	0.553 (0.002)	0.539 (0.005)	0.486 (0.011)	0.441 (0.016)	0.516 (0.009)
G	0.607 (0.001)	0.601 (0.002)	0.585 (0.006)	0.534 (0.012)	0.481 (0.019)	0.562 (0.009)
NS	0.542 (0.045)	0.524 (0.048)	0.491 (0.050)	0.412 (0.054)	0.343 (0.051)	0.463 (0.027)
RC	0.541 (0.045)	0.525 (0.048)	0.492 (0.051)	0.414 (0.055)	0.345 (0.053)	0.463 (0.027)
MA	0.544 (0.044)	0.534 (0.047)	0.521 (0.048)	0.471 (0.051)	0.416 (0.053)	0.497 (0.022)
NG	0.538 (0.046)	0.533 (0.049)	0.523 (0.050)	0.476 (0.053)	0.425 (0.055)	0.499 (0.022)
SD	0.545 (0.044)	0.536 (0.046)	0.527 (0.048)	0.489 (0.048)	0.451 (0.050)	0.509 (0.020)
AR	0.542 (0.045)	0.529 (0.050)	0.518 (0.051)	0.478 (0.050)	0.443 (0.054)	0.502 (0.021)
GK	0.543 (0.046)	0.532 (0.051)	0.524 (0.052)	0.481 (0.050)	0.443 (0.057)	0.505 (0.022)

Supplementary Table S6: Marginal selection accuracy (top 10%) and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
NS	0.366 (0.040)	0.491 (0.034)	0.531 (0.038)
RC	0.365 (0.040)	0.492 (0.034)	0.533 (0.037)
MA	0.403 (0.027)	0.521 (0.022)	0.567 (0.023)
NG	0.402 (0.025)	0.525 (0.020)	0.571 (0.020)
SD	0.419 (0.019)	0.531 (0.017)	0.578 (0.016)
AR	0.405 (0.021)	0.528 (0.017)	0.573 (0.017)
GK	0.407 (0.021)	0.527 (0.019)	0.580 (0.017)

Supplementary Table S7: Marginal bias of genetic variances and standard error by delta ratios and treatments.

	9:1	3:1	1:1	1:3	1:9	AVG
I	-0.067 (0.004)	-0.059 (0.003)	-0.054 (0.004)	-0.038 (0.010)	-0.032 (0.010)	-0.050 (0.004)
A	-0.009 (0.002)	-0.002 (0.002)	-0.006 (0.002)	-0.001 (0.004)	-0.005 (0.005)	-0.005 (0.001)
G	-0.017 (0.002)	-0.012 (0.002)	-0.015 (0.003)	-0.008 (0.004)	-0.010 (0.004)	-0.012 (0.001)
NS	-0.029 (0.016)	-0.025 (0.017)	-0.033 (0.017)	-0.031 (0.018)	-0.033 (0.015)	-0.030 (0.006)
RC	-0.033 (0.019)	-0.028 (0.019)	-0.036 (0.018)	-0.036 (0.020)	-0.038 (0.016)	-0.034 (0.007)
MA	-0.028 (0.016)	-0.022 (0.015)	-0.022 (0.011)	-0.013 (0.010)	-0.014 (0.007)	-0.020 (0.005)
NG	-0.041 (0.024)	-0.028 (0.021)	-0.025 (0.015)	-0.014 (0.012)	-0.012 (0.007)	-0.024 (0.007)
SD	-0.021 (0.016)	-0.016 (0.014)	-0.011 (0.013)	0.008 (0.006)	0.008 (0.003)	-0.007 (0.005)
AR	-0.031 (0.018)	-0.025 (0.018)	-0.023 (0.014)	-0.008 (0.010)	-0.002 (0.006)	-0.018 (0.006)
GK	-0.033 (0.020)	-0.025 (0.020)	-0.023 (0.015)	-0.016 (0.007)	-0.017 (0.007)	-0.023 (0.006)

Supplementary Table S8: Marginal bias of genetic variances and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
NS	-0.063 (0.002)	-0.010 (0.002)	-0.017 (0.002)
RC	-0.070 (0.002)	-0.012 (0.003)	-0.020 (0.002)
MA	-0.043 (0.006)	-0.005 (0.002)	-0.012 (0.002)
NG	-0.055 (0.011)	-0.004 (0.002)	-0.013 (0.003)
SD	-0.027 (0.012)	0.006 (0.004)	0.001 (0.003)
AR	-0.043 (0.010)	0.001 (0.004)	-0.012 (0.002)
GK	-0.050 (0.009)	-0.007 (0.002)	-0.012 (0.002)

Supplementary Table S9: Marginal bias of spatial variances and standard error by delta ratios and treatments

	9:1	3:1	1:1	1:3	1:9	AVG
I	-0.033 (0.013)	-0.082 (0.026)	-0.119 (0.048)	-0.208 (0.089)	-0.288 (0.128)	-0.146 (0.036)
A	-0.033 (0.011)	-0.056 (0.022)	-0.111 (0.050)	-0.214 (0.088)	-0.290 (0.130)	-0.141 (0.036)
G	-0.023 (0.015)	-0.054 (0.026)	-0.101 (0.054)	-0.205 (0.092)	-0.289 (0.131)	-0.134 (0.037)
RC	-0.052 (0.000)	-0.138 (0.001)	-0.309 (0.002)	-0.553 (0.002)	-0.761 (0.002)	-0.362 (0.070)
NG	-0.027 (0.001)	-0.038 (0.003)	-0.080 (0.002)	-0.178 (0.003)	-0.275 (0.003)	-0.120 (0.025)
SD	0.017 (0.007)	-0.004 (0.007)	-0.027 (0.010)	-0.040 (0.016)	0.035 (0.009)	-0.004 (0.008)
AR	-0.051 (0.000)	-0.083 (0.019)	-0.059 (0.014)	-0.112 (0.005)	-0.200 (0.005)	-0.101 (0.015)
GK	-0.036 (0.011)	-0.055 (0.030)	-0.076 (0.006)	-0.161 (0.002)	-0.244 (0.003)	-0.114 (0.021)

Supplementary Table S10: Marginal bias of spatial variances and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
RC	-0.362 (0.131)	-0.363 (0.132)	-0.362 (0.131)
NG	-0.115 (0.047)	-0.121 (0.047)	-0.123 (0.047)
SD	-0.014 (0.014)	-0.013 (0.016)	0.015 (0.011)
AR	-0.109 (0.023)	-0.095 (0.029)	-0.099 (0.031)
GK	-0.129 (0.032)	-0.111 (0.041)	-0.103 (0.045)

Supplementary Table S11: Marginal bias of residual variances and standard error by delta ratios and treatments.

	9:1	3:1	1:1	1:3	1:9	AVG
I	0.089 (0.010)	0.132 (0.023)	0.167 (0.052)	0.233 (0.098)	0.303 (0.132)	0.185 (0.035)
A	0.040 (0.008)	0.063 (0.021)	0.119 (0.051)	0.207 (0.091)	0.282 (0.125)	0.142 (0.035)
G	0.037 (0.009)	0.066 (0.023)	0.115 (0.053)	0.206 (0.091)	0.282 (0.125)	0.141 (0.035)
NS	0.069 (0.015)	0.153 (0.016)	0.322 (0.017)	0.561 (0.018)	0.765 (0.016)	0.374 (0.069)
RC	0.073 (0.017)	0.155 (0.017)	0.325 (0.018)	0.564 (0.019)	0.766 (0.016)	0.377 (0.069)
MA	0.056 (0.014)	0.092 (0.014)	0.148 (0.011)	0.213 (0.008)	0.268 (0.006)	0.155 (0.021)
NG	0.054 (0.020)	0.047 (0.015)	0.058 (0.013)	0.093 (0.009)	0.127 (0.005)	0.076 (0.009)
SD	0.009 (0.012)	0.002 (0.012)	-0.001 (0.011)	0.004 (0.002)	0.021 (0.001)	0.007 (0.004)
AR	0.071 (0.016)	0.097 (0.035)	0.042 (0.028)	0.015 (0.007)	0.008 (0.003)	0.046 (0.012)
GK	0.056 (0.027)	0.062 (0.051)	0.041 (0.021)	0.055 (0.006)	0.069 (0.005)	0.057 (0.011)

Supplementary Table S12: Marginal bias of residual variances and standard error by relationship and spatial methods.

	Identity (I)	Pedigree (A)	Genomic (G)
NS	0.406 (0.129)	0.358 (0.129)	0.358 (0.128)
RC	0.411 (0.128)	0.359 (0.129)	0.359 (0.128)
MA	0.177 (0.036)	0.144 (0.040)	0.145 (0.040)
NG	0.101 (0.011)	0.062 (0.018)	0.065 (0.017)
SD	0.021 (0.005)	0.004 (0.005)	-0.003 (0.007)
AR	0.080 (0.028)	0.026 (0.011)	0.033 (0.014)
GK	0.099 (0.017)	0.042 (0.010)	0.029 (0.010)

Supplementary Text T1: Factors influencing accuracy within- and across-locations

The study focused on simulated data where accuracy can be computed directly, as the true values of the coefficients are known. However, for completeness, we provide here an alternative framework where the accuracy can be estimated if variance components are known, and the statistical model is identical to the real model. We also provide generalizations to multi-environmental trials.

Single-environmental trial: As the true statistical model, consider a linear model where phenotypes are modeled as a function of controllable factors (Xb), genetics (Zu), spatial variation (s), and residuals (e):

$$\begin{aligned} &y = Xb + Zu + s + e \\ &y \sim N(Xb, V) \\ &V = G + S + R = ZKZ'\sigma_g^2 + H\sigma_s^2 + I\sigma_e^2 \\ &\mathrm{cov}(G, S) = \mathrm{cov}(G, R) = \mathrm{cov}(S, R) = 0 \\ &u \sim N\left(0, K\sigma_g^2\right) \\ &s \sim N(0, H\sigma_s^2) \\ &I \sim N(0, I\sigma_e^2) \end{aligned}$$

Where y is the vector of phenotypes, X is a matrix of fixed effects that capture the overall mean and other known nuisance parameters, b is the vector of fixed effects, Z is the design matrix of individuals, u is the vector of genetic values with covariance $K\sigma_g^2$, s is the vector of spatial effects with covariance $H\sigma_s^2$, and e is the vector of i.i.d. residuals. Since Xb and s characterize nuisance parameters when genetics is the factor of interest, the conditional distribution of y with respect to fixed and spatial effects is given by:

$$\begin{aligned} &\tilde{y} = y - Xb - s \\ &\tilde{y} \sim N\left(0, \tilde{V}\right) \\ &\tilde{V} = G + R = ZKZ'\sigma_g^2 + I\sigma_e^2 \end{aligned}$$

The accuracy of genetic coefficients is defined by the correlation between estimated and true values.

$$A = \mathrm{cor}(\hat{g}, g) = \frac{\mathrm{cov}(\hat{g}, g)}{\sqrt{\mathrm{var}(\hat{g})\mathrm{var}(g)}}$$

Under the assumption that the statistical model corresponds to the true model, $\mathrm{var}(g) = \mathrm{var}(\hat{g})$, and the variance components are known, a deterministic calculation of accuracy is derived as follows:

$$A = \frac{\mathrm{cov}(\hat{g}, g)}{\mathrm{var}(\hat{g})} = \sqrt{\frac{G\tilde{V}^{-1}G}{G}} = \sqrt{G^{-1}G\tilde{V}^{-1}G} = \sqrt{\tilde{V}^{-1}G} = \sqrt{\left(ZKZ'\sigma_g^2 + I\sigma_e^2\right)^{-1}ZKZ'\sigma_g^2}$$

Where the diagonal elements (a_{ii}) correspond to the genetic accuracy of each observation. Also, note that $V^{-1}G$ is the matrix generalization of the plot-level heritability. Thus, the accuracy of the selection of non-replicated trials are generally simplified as the squared root of plot-level heritability ($a = \sqrt{h^2}$). The entry-level accuracy of each genotypic value is expressed as:

$$A_{g_i} = Z_i'AZ_i = \sqrt{\frac{Z_i'KZ_i\sigma_g^2}{Z_i'(G + R)^{-1}Z_i}}$$

For illustration purposes, let $K = I$, such that the formulation is simplified into:

$$A_{g_i} = \sqrt{\frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{n_i}}}$$

Where n corresponds to the number of replicates (or, in matrix notation, $n_i = Z_i'Z_i$). By dismissing the genetic relationship matrix, such simplification does not work for predictions ($n = 0$), however, it illustrates how the variance components and the number of replications influence accuracy.

Implications for spatial designs: The key implication of correctly estimating and capture field patterns are associated with increasing the accuracy by increasing the heritability. Spatial variance not captured by the spatial term is passed to the residuals. The residual variance has an additive contribution to $\tilde{V}$ and, therefore, a negative impact on accuracy ($A = \sqrt{\tilde{V}^{-1}G}$).

Implications under P-REP designs: In an experimental setting where genotypes are unbalanced, the entry-level accuracy of any given individual is a function of the amount of phenotypic variance explained by genetics ($V^{-1}G \approx \sigma_g^2 \div \sigma_y^2$) and the number of replicated (n_i), as clearly depicted in the simplified formulation described above. In addition, the entry-level accuracy will be a function of the relationship between the target individual and all other entries through the information captured by the genetic relationship matrix K, since related individuals serve as partial replication.

Multi-environmental trials with GxE interaction: When multiple environments are modeled jointly, consider the model:

$$y = Xb + Zu + s + e$$
$$y \sim N(Xb, V)$$
$$V = G + S + R = Z(K \otimes \Sigma_g)Z' + H \otimes \Sigma_s + I \otimes \Sigma_e$$

Where the variance components are generalized into matrices accommodate location-specific ($\Sigma_g, \Sigma_s, \Sigma_e$). The genetic covariate Σ_g contain the variance of each location in the diagonal and the genetic covariance between each pair of locations in the off diagonals. Covariances Σ_s and Σ_e are diagonal matrices since the residuals and spatial patterns are not correlated across environments. The multivariate model preserves the same adjusted model for nuisance parameters:

$$\tilde{y} = y - Xb - s$$
$$y \sim N(0, \tilde{V})$$
$$\tilde{V} = G + R$$

An adequate depiction of $\tilde{V}$ entails the proper conditioning of the phenotype regarding the spatial variation. The importance of capturing the spatial variation across environments remains since variation not captured by the spatial term within a given environment will pass on to the residual variance and, consequently, affect the heritability of that location.

The accuracy across environments operated as a selection index, where the multiple environments may or may not carry the same weight for selection. Following the Smith-Hazel index, the entry-basis accuracy can be described as:

$$A_{g_i} = \sqrt{\frac{G_{i,y}\tilde{V}^{-1}G_{y,i}}{G_{ii}}}$$

As $G_{i,y}$ is a vector, the covariance between the i^{th} individual and the observed values defined by:

$$G_{i,y} = Z_i'K \otimes \Sigma_{g,TPE}$$

The vector $\Sigma_{g,TPE}$ contain the covariances between each environment and the target population of environments (TPE). If environments are weighted according to the amount of genetic variance estimated from each environment, then $\Sigma_{g,TPE} = (\sigma_{g1}, \sigma_{g2}, \ldots, \sigma_{g2})$.

Thus, in addition to the variance components and relationship, the multi-environmental accuracy will rely on the covariance components (Σ_g) which inform the information flow among environments. The variance components corresponding to residuals and spatial have an inversely proportional impact on the accuracy (V^{-1}) and the contribution of each environment towards the across-environment accuracy.

Implications under P-LOC designs: Both plot-level and entry-level accuracies are affected by P-LOC designs, where not all individuals are observed across all locations. The genetic merit of genotypes not observed within a given environment will be a prediction based on the relationship. As aforementioned, the genetic merit of non-observed individuals displays low plot-level accuracy since $n = 0$. Conversely, the entry-level accuracy is less sensitive to missing information due to the information being leveraged from a genetic relationship where the individual is not observed, as well as from the information provided by the locations where the genotype is observed.

In general, the plot-level accuracy is a function of three factors: 1) the heritability of the target location E, approximate as the mean diagonal of $\left(K \otimes \Sigma_{gE} + I \otimes \Sigma_{eE}\right)^{-1} K \otimes \Sigma_{gE}$; 2) the genetic relationship (K) with the individuals observed in the location E; and 3) any information gained from correlated environments, based on the genetic covariance Σ_g, including environments where the given genotype is observed. The entry-level accuracy, as a summary of the genetic merit over the TPE, will be a function of the heritability across multiple locations, the relationship among individuals, and the number of times the individual was observed, and the genetic covariance among environments.