Single-step genomic selection: accommodating several key issues

Aidan McGarty[1]
Brian Cullis[1], Ahsan Asif[2] and Kristy Hobson[2]

September 2, 2024

Mixed Models and Experiment Design Lab (MMaED)[1]
National Institute for Applied Statistics Research Australia
University of Wollongong
`amcgarty@uow.edu.au`

Chickpea Breeding Australia[2]
NSW Department of Primary Industries | Agriculture

Overview

Overview

- **Genomic selection**

Overview

- Genomic selection
- Motivating Example

Overview

- **Genomic selection**
- **Motivating Example**
- **Complexities**

Genomic Selection - What is it?

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:
  - Genes affecting a relevant trait have an association to a set of known genetic covariates (markers)

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:
  - Genes affecting a relevant trait have an association to a set of known genetic covariates (markers)
  - These associations are then exploited to determine an individuals expression of this trait given the covariate values are known for an individual

Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:
    - Genes affecting a relevant trait have an association to a set of known genetic covariates (markers)
    - These associations are then exploited to determine an individuals expression of this trait given the covariate values are known for an individual
- In practice:

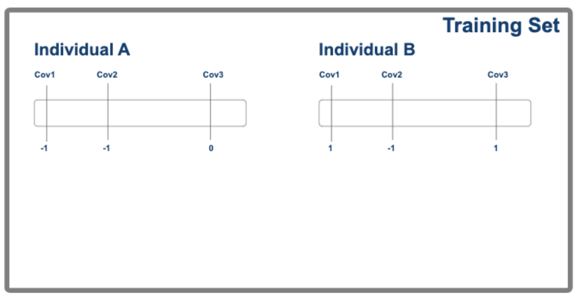
Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:
  - Genes affecting a relevant trait have an association to a set of known genetic covariates (markers)
  - These associations are then exploited to determine an individuals expression of this trait given the covariate values are known for an individual
- In practice:
  - A training dataset is used to estimate the relationship between the relevant trait and the known genetic covariates

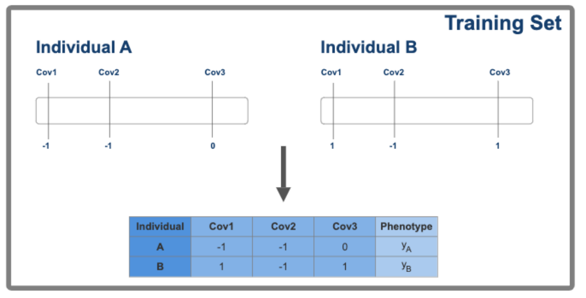
Genomic Selection - What is it?

- Plant breeding programs aim to develop superior varieties for commercial release
- This requires extensive evaluation experiments to identify superior material
- Genomic selection (GS) is a cost and time saving application of statistical and genetic theory whereby selections can be made for a set of lines without evaluation experiments
- Statistical principle:
    - Genes affecting a relevant trait have an association to a set of known genetic covariates (markers)
    - These associations are then exploited to determine an individuals expression of this trait given the covariate values are known for an individual
- In practice:
    - A training dataset is used to estimate the relationship between the relevant trait and the known genetic covariates
    - These estimated relationships are then used to predict the performance of a set of test individuals

Genomic Selection

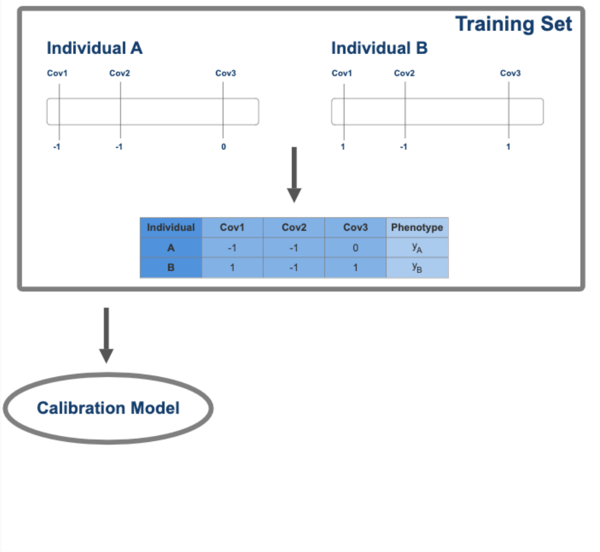


Genomic Selection



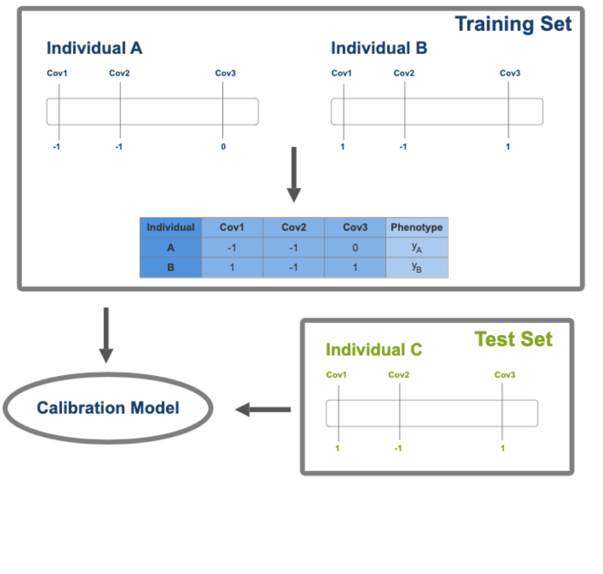
Individual	Cov1	Cov2	Cov3	Phenotype
A	-1	-1	0	y_A
B	1	-1	1	y_B

Genomic Selection



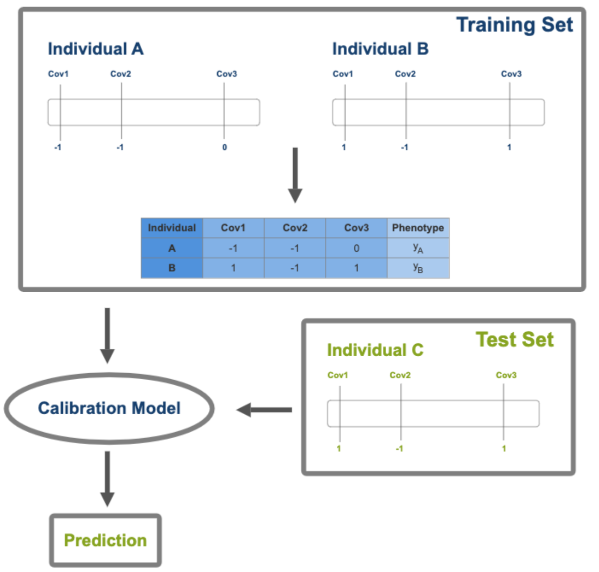
Individual	Cov1	Cov2	Cov3	Phenotype
A	-1	-1	0	y_A
B	1	-1	1	y_B

Genomic Selection



Individual	Cov1	Cov2	Cov3	Phenotype
A	-1	-1	0	y_A
B	1	-1	1	y_B

Genomic Selection



Individual	Cov1	Cov2	Cov3	Phenotype
A	-1	-1	0	y_A
B	1	-1	1	y_B

Calibration Model

Calibration Model

- We choose a single step factor analytic linear mixed model (FALMM) (Smith et al. [2024]) as the calibration model:

Calibration Model

- We choose a single step factor analytic linear mixed model (FALMM) (Smith et al. [2024]) as the calibration model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

Calibration Model

- We choose a single step factor analytic linear mixed model (FALMM) (Smith et al. [2024]) as the calibration model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- The key component of this model for the purpose of this talk are the variety effects nested within environments (VE effects) ($\mathbf{u}_g$) which are partitioned into an additive and non-additive genetic component; $\mathbf{u}_g = \mathbf{u}_a + \mathbf{u}_e$

Calibration Model

- We choose a single step factor analytic linear mixed model (FALMM) (Smith et al. [2024]) as the calibration model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- The key component of this model for the purpose of this talk are the variety effects nested within environments (VE effects) ($\mathbf{u}_g$) which are partitioned into an additive and non-additive genetic component; $\mathbf{u}_g = \mathbf{u}_a + \mathbf{u}_e$
- We model both $\mathbf{u}_a$ and $\mathbf{u}_e$ with factor analytic structures, where the covariance of these variety effects between environments is modelled by k_a and k_e unknown factors, respectively

Calibration Model

- We choose a single step factor analytic linear mixed model (FALMM) (Smith et al. [2024]) as the calibration model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- The key component of this model for the purpose of this talk are the variety effects nested within environments (VE effects) ($\mathbf{u}_g$) which are partitioned into an additive and non-additive genetic component; $\mathbf{u}_g = \mathbf{u}_a + \mathbf{u}_e$
- We model both $\mathbf{u}_a$ and $\mathbf{u}_e$ with factor analytic structures, where the covariance of these variety effects between environments is modelled by k_a and k_e unknown factors, respectively
- For the additive VE effects $\mathbf{u}_a$, covariance between varieties is also modelled via a known relationship matrix denoted $\mathbf{G}$

Calibration Model - Genetic relatedness

Calibration Model - Genetic relatedness

- This relationship matrix **G** may be formed via ancestral records, marker information or a combination of both (more on this later)

Calibration Model - Genetic relatedness

- This relationship matrix **G** may be formed via ancestral records, marker information or a combination of both (more on this later)
- The matrix formed via marker scores (genetic covariates) is known as the genomic relationship matrix (GRM), often denoted $\mathbf{K} = \mathbf{M}\mathbf{M}^T$ where **M** is the matrix of centered and scaled marker scores (VanRaden [2008])

Calibration Model - Genetic relatedness

- This relationship matrix **G** may be formed via ancestral records, marker information or a combination of both (more on this later)
- The matrix formed via marker scores (genetic covariates) is known as the genomic relationship matrix (GRM), often denoted $\mathbf{K} = \mathbf{M}\mathbf{M}^T$ where **M** is the matrix of centered and scaled marker scores (VanRaden [2008])
- The matrix formed via ancestral records is denoted **A** and known as the numerator relationship matrix (NRM) (Oakey et al. [2007])

Calibration Model - Covariance between environments

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

- Where $\boldsymbol{\Lambda}_a = \left[\boldsymbol{\lambda}_{a_1}, \ldots, \boldsymbol{\lambda}_{a_{k_a}}\right]$ is the $(p \times k_a)$ matrix of loadings with $\boldsymbol{\lambda}_{a_i}$ the p-vector of loadings for the ith factor (note p is the number of environments).

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

- Where $\boldsymbol{\Lambda}_a = \left[\boldsymbol{\lambda}_{a_1}, \ldots, \boldsymbol{\lambda}_{a_{k_a}}\right]$ is the $(p \times k_a)$ matrix of loadings with $\boldsymbol{\lambda}_{a_i}$ the p-vector of loadings for the ith factor (note p is the number of environments).
- $\mathbf{I}_v$ is an identity matrix of dimension $(v \times v)$, v being the number of varieties and $\mathbf{f}_a$ and $\boldsymbol{\delta}_a$ are the vk_a-vector of variety scores and vp-vector of lack-of-fit effects respectively with the following joint distribution:

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

- Where $\boldsymbol{\Lambda}_a = \left[\boldsymbol{\lambda}_{a_1}, \ldots, \boldsymbol{\lambda}_{a_{k_a}}\right]$ is the $(p \times k_a)$ matrix of loadings with $\boldsymbol{\lambda}_{a_i}$ the p-vector of loadings for the ith factor (note p is the number of environments).
- $\mathbf{I}_v$ is an identity matrix of dimension $(v \times v)$, v being the number of varieties and $\mathbf{f}_a$ and $\boldsymbol{\delta}_a$ are the vk_a-vector of variety scores and vp-vector of lack-of-fit effects respectively with the following joint distribution:

$$\begin{bmatrix} \mathbf{f}_a \\ \boldsymbol{\delta}_a \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{D}_a \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \psi_a \otimes \mathbf{G} \end{bmatrix}\right)$$

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

- Where $\boldsymbol{\Lambda}_a = \left[\boldsymbol{\lambda}_{a_1}, \ldots, \boldsymbol{\lambda}_{a_{k_a}}\right]$ is the $(p \times k_a)$ matrix of loadings with $\boldsymbol{\lambda}_{a_i}$ the p-vector of loadings for the ith factor (note p is the number of environments).
- $\mathbf{I}_v$ is an identity matrix of dimension $(v \times v)$, v being the number of varieties and $\mathbf{f}_a$ and $\boldsymbol{\delta}_a$ are the vk_a-vector of variety scores and vp-vector of lack-of-fit effects respectively with the following joint distribution:

$$\begin{bmatrix} \mathbf{f}_a \\ \boldsymbol{\delta}_a \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{D}_a \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \psi_a \otimes \mathbf{G} \end{bmatrix}\right)$$

- Note $\boldsymbol{\Lambda}_a$ has been rotated to the principal components solution and as such $\mathbf{D}_a$ is a matrix with the eigenvalues of $\boldsymbol{\Lambda}_a^*$ on the diagonals ($\boldsymbol{\Lambda}_a^*$ is the unrotated loadings)

Calibration Model - Covariance between environments

- This factor analytic structure for the additive component $\mathbf{u}_a$ is of the form:

$$\mathbf{u}_a = (\boldsymbol{\Lambda}_a \otimes \mathbf{I}_v)\mathbf{f}_a + \boldsymbol{\delta}_a$$

- Where $\boldsymbol{\Lambda}_a = \left[\boldsymbol{\lambda}_{a_1}, \ldots, \boldsymbol{\lambda}_{a_{k_a}}\right]$ is the $(p \times k_a)$ matrix of loadings with $\boldsymbol{\lambda}_{a_i}$ the p-vector of loadings for the ith factor (note p is the number of environments).
- $\mathbf{I}_v$ is an identity matrix of dimension $(v \times v)$, v being the number of varieties and $\mathbf{f}_a$ and $\boldsymbol{\delta}_a$ are the vk_a-vector of variety scores and vp-vector of lack-of-fit effects respectively with the following joint distribution:

$$\begin{bmatrix} \mathbf{f}_a \\ \boldsymbol{\delta}_a \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{D}_a \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \psi_a \otimes \mathbf{G} \end{bmatrix}\right)$$

- Note $\boldsymbol{\Lambda}_a$ has been rotated to the principal components solution and as such $\mathbf{D}_a$ is a matrix with the eigenvalues of $\boldsymbol{\Lambda}_a^*$ on the diagonals ($\boldsymbol{\Lambda}_a^*$ is the unrotated loadings)
- ψ_a is a diagonal matrix with the site specific variances ψ_{a_i} on the diagonals

Motivating example

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)
- Including $v = 14629$ varieties in the training data set (i.e both phenotyped and genotyped)

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)
- Including $v = 14629$ varieties in the training data set (i.e both phenotyped and genotyped)
- A total of $m = 4303$ retained markers (genetic covariates)

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)
- Including $v = 14629$ varieties in the training data set (i.e both phenotyped and genotyped)
- A total of $m = 4303$ retained markers (genetic covariates)
- Across $p = 11$ environments including field, glass house and semi-controlled shade house experiments

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)
- Including $v = 14629$ varieties in the training data set (i.e both phenotyped and genotyped)
- A total of $m = 4303$ retained markers (genetic covariates)
- Across $p = 11$ environments including field, glass house and semi-controlled shade house experiments
- Aim is to predict disease resistance among a test set of unphenotyped individuals

Motivating example

- Australian Chickpea disease resistance multi-environment trial (MET)
- Including $v = 14629$ varieties in the training data set (i.e both phenotyped and genotyped)
- A total of $m = 4303$ retained markers (genetic covariates)
- Across $p = 11$ environments including field, glass house and semi-controlled shade house experiments
- Aim is to predict disease resistance among a test set of unphenotyped individuals
- The final model fit included $k_a = 2$ additive and $k_e = 1$ non-additive factors

Motivating example

Complexities

Complexities

1. Variety by environment interaction (GxE)

Complexities

1. Variety by environment interaction (GxE)
2. Complex trial designs:

Complexities

1. Variety by environment interaction (GxE)
2. Complex trial designs:
   - Need to accommodate a range of complex trial designs including both field and lab experiments

Complexities

1. Variety by environment interaction (GxE)
2. Complex trial designs:
    - Need to accommodate a range of complex trial designs including both field and lab experiments
3. **Linear dependencies in matrix of marker scores:**

Complexities

1. Variety by environment interaction (GxE)
2. Complex trial designs:
    - Need to accommodate a range of complex trial designs including both field and lab experiments
3. **Linear dependencies in matrix of marker scores:**
    - Collinear covariates

Complexities

1. Variety by environment interaction (GxE)
2. Complex trial designs:
   - Need to accommodate a range of complex trial designs including both field and lab experiments
3. **Linear dependencies in matrix of marker scores:**
   - Collinear covariates
   - More individuals than markers

Linear dependencies in matrix of marker scores

Linear dependencies in matrix of marker scores

- Results in a singular and hence non-invertible genomic relationship matrix

Linear dependencies in matrix of marker scores

- Results in a singular and hence non-invertible genomic relationship matrix
- This inverse is required when solving the mixed model equations (MMEs)

Linear dependencies in matrix of marker scores

- Results in a singular and hence non-invertible genomic relationship matrix
- This inverse is required when solving the mixed model equations (MMEs)
- Several strategies exist to overcome this complexity however the method we have used is a blending of the **K** and **A** relationship matrices

Blended genetic variance matrix

Blended genetic variance matrix

- Combining genomic information with pedigree information has been suggested by several authors (Legarra et al. [2009],Vitezica et al. [2011], Meyer et al. [2018])

Blended genetic variance matrix

- Combining genomic information with pedigree information has been suggested by several authors (Legarra et al. [2009],Vitezica et al. [2011], Meyer et al. [2018])
- This will always result in an invertible genetic variance matrix even if the GRM $\mathbf{K}$ is non-invertible as a positive definite matrix (as is guaranteed when forming $\mathbf{A}$) plus a positive semi definite matrix will be positive definite (Harville [1997])

Blended genetic variance matrix

- Combining genomic information with pedigree information has been suggested by several authors (Legarra et al. [2009],Vitezica et al. [2011], Meyer et al. [2018])
- This will always result in an invertible genetic variance matrix even if the GRM **K** is non-invertible as a positive definite matrix (as is guaranteed when forming **A**) plus a positive semi definite matrix will be positive definite (Harville [1997])
- We specify **G** as:

$$\mathbf{G} = \lambda \mathbf{K} + (1 - \lambda)\mathbf{A}$$

Blended genetic variance matrix

- Combining genomic information with pedigree information has been suggested by several authors (Legarra et al. [2009],Vitezica et al. [2011], Meyer et al. [2018])
- This will always result in an invertible genetic variance matrix even if the GRM **K** is non-invertible as a positive definite matrix (as is guaranteed when forming **A**) plus a positive semi definite matrix will be positive definite (Harville [1997])
- We specify **G** as:

$$\mathbf{G} = \lambda\mathbf{K} + (1-\lambda)\mathbf{A}$$

- The added parameter λ controls the weighting from the two sources of relatedness, with genomic information dominating when $\lambda \approx 1$ and pedigree information dominating when $\lambda \approx 0$

Choice of λ

Choice of λ

- We choose the value of λ which maximises the residual loglikelihood

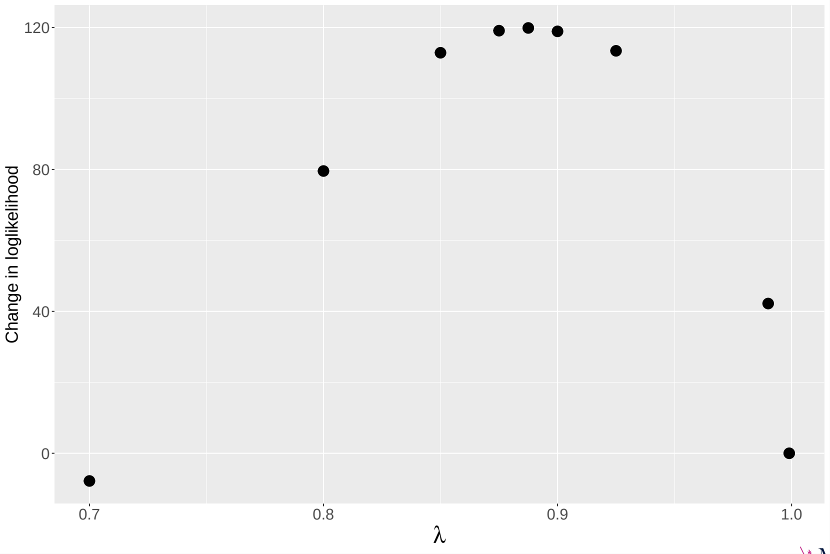
Choice of λ

- We choose the value of λ which maximises the residual loglikelihood
- This is done via maximising the full residual likelihood where λ is simply an extra parameter to estimate

Choice of λ

- We choose the value of λ which maximises the residual loglikelihood
- This is done via maximising the full residual likelihood where λ is simply an extra parameter to estimate
- However this can be shown graphically through profiling on the parameter of interest (λ) and assessing the residual log-likelihood under a range of values for this parameter

Choice of λ



Predictions

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information
- Considering $\mathbf{f}_a$ has the following variance, where $\mathbf{f}_a$ is reordered as environments within varieties:

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information
- Considering $\mathbf{f}_a$ has the following variance, where $\mathbf{f}_a$ is reordered as environments within varieties:

$$\text{var}\begin{bmatrix}\mathbf{f}_{a_1}\\ \mathbf{f}_{a_2}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix} \otimes \mathbf{D}_a$$

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information
- Considering $\mathbf{f}_a$ has the following variance, where $\mathbf{f}_a$ is reordered as environments within varieties:

$$\text{var}\begin{bmatrix} \mathbf{f}_{a_1} \\ \mathbf{f}_{a_2} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} \otimes \mathbf{D}_a$$

- Where $\mathbf{f}_{a_1}$ and $\mathbf{G}_{11}$ correspond to the partition of $\mathbf{f}$ and $\mathbf{G}$ relating to varieties with phenotypic records (training set) and $\mathbf{f}_{a_2}$ and $\mathbf{G}_{22}$ corresponds to the partitions relating to varieties without phenotypic records (test set)

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information
- Considering $\mathbf{f}_a$ has the following variance, where $\mathbf{f}_a$ is reordered as environments within varieties:

$$\text{var}\begin{bmatrix}\mathbf{f}_{a_1}\\ \mathbf{f}_{a_2}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix} \otimes \mathbf{D}_a$$

- Where $\mathbf{f}_{a_1}$ and $\mathbf{G}_{11}$ correspond to the partition of $\mathbf{f}$ and $\mathbf{G}$ relating to varieties with phenotypic records (training set) and $\mathbf{f}_{a_2}$ and $\mathbf{G}_{22}$ corresponds to the partitions relating to varieties without phenotypic records (test set)
- We then form predicted values for the test set:

Predictions

- The model described then allows us to form predictions of variety scores for all varieties in the test set, that is, those with genotypic information but without phenotypic information
- Considering $\mathbf{f}_a$ has the following variance, where $\mathbf{f}_a$ is reordered as environments within varieties:

$$\text{var}\begin{bmatrix}\mathbf{f}_{a_1}\\ \mathbf{f}_{a_2}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix} \otimes \mathbf{D}_a$$

- Where $\mathbf{f}_{a_1}$ and $\mathbf{G}_{11}$ correspond to the partition of $\mathbf{f}$ and $\mathbf{G}$ relating to varieties with phenotypic records (training set) and $\mathbf{f}_{a_2}$ and $\mathbf{G}_{22}$ corresponds to the partitions relating to varieties without phenotypic records (test set)
- We then form predicted values for the test set:

$$\tilde{\mathbf{f}}_{a_2} = (\mathbf{G}_{21}\mathbf{G}_{11}^{-1} \otimes \mathbf{I}_{k_a})\tilde{\mathbf{f}}_{a_1}$$

Conclusion

Conclusion

- Genomic selection is a powerful tool which has the capacity to save significant time and money in plant breeding programs

Conclusion

- Genomic selection is a powerful tool which has the capacity to save significant time and money in plant breeding programs
- However many complexities arise in implementation, a few of which have been explored here, specifically singularities in the GRM ($\mathbf{K}$)

Conclusion

- Genomic selection is a powerful tool which has the capacity to save significant time and money in plant breeding programs
- However many complexities arise in implementation, a few of which have been explored here, specifically singularities in the GRM ($\mathbf{K}$)
- Combining both ancestral and genomic relationship matrices both overcomes this issue

Conclusion

- Genomic selection is a powerful tool which has the capacity to save significant time and money in plant breeding programs
- However many complexities arise in implementation, a few of which have been explored here, specifically singularities in the GRM ($\mathbf{K}$)
- Combining both ancestral and genomic relationship matrices both overcomes this issue
- Finally DWReml (Butler, pers comm) can handle the high computational demand when fitting complex single-step factor models which include genetic relatedness such as the motivating example

References i

A Asif, A McGarty, B Cullis, and K Hobson. Identification of resistance to aggressive ascochyta rabiei isolates in australian chickpea breeding lines using genomic selection [in prep]. 2024.

D A Harville. *Matrix algebra from a statisticians perspective*. Springer-Verlag, New York, 1997.

A. Legarra, I. Aguilar, and I. Misztal. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92:4656–4663, 9 2009. ISSN 00220302. doi: 10.3168/jds.2009-2061.

Karin Meyer, Bruce Tier, and Andrew Swan. Estimates of genetic trend for single step genomic evaluations. *Genetics Selection Evolution*, 50:39, 12 2018. ISSN 1297-9686. doi: 10.1186/s12711-018-0410-1.

H Oakey, A P Verbyla, B R Cullis, X Wei, and W S Pitchford. Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics*, 114, 2007. ISSN 00405752. doi: 10.1007/s00122-007-0515-3.

A B Smith, A S K Shunmugam, D G Butler, and B R Cullis. Plant variety selection using interaction classes derived from factor analytic linear mixed models: models with information on genetic relatedness. *Journal of Agricultural Science [under revision]*, 2024.

References ii

P M VanRaden. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91:4414–4423, 2008. ISSN 00220302. doi: 10.3168/jds.2007-0980. URL `http://linkinghub.elsevier.com/retrieve/pii/S0022030208709901`.

Z. G. Vitezica, I. Aguilar, I. Misztal, and A. Legarra. Bias in genomic predictions for populations under selection. *Genetics Research*, 93:357–366, 10 2011. ISSN 0016-6723. doi: 10.1017/S001667231100022X.