Multi-environment trial analysis of count data with complex variance structures using generalised linear mixed models

Michael Mumford Dr. Alan Huang Dr. Alison Kelly

November 25, 2025

BILL & MELINDA GATES foundation
EIAR Ethiopian Institute of Agricultural Research
THE UNIVERSITY OF QUEENSLAND AUSTRALIA
QAAFI Queensland Alliance for Agriculture and Food Innovation
Queensland Government

Outline

- Motivating multi-environment trial (MET) data

Outline

- Motivating multi-environment trial (MET) data
- Mixed models

Outline

- Motivating multi-environment trial (MET) data
- Mixed models
- Mean parameterised Conway-Maxwell Poisson (CMP_{μ}) distribution

Outline

- Motivating multi-environment trial (MET) data
- Mixed models
- Mean parameterised Conway-Maxwell Poisson (CMP_{μ}) distribution
- Statistical model

Outline

- Motivating multi-environment trial (MET) data
- Mixed models
- Mean parameterised Conway-Maxwell Poisson (CMP_{μ}) distribution
- Statistical model
  - Factor-analytic Conway-Maxwell Poisson (FA-CMP_{μ}) model

Outline

- Motivating multi-environment trial (MET) data
- Mixed models
- Mean parameterised Conway-Maxwell Poisson (CMP_μ) distribution
- Statistical model
  - Factor-analytic Conway-Maxwell Poisson (FA-CMP_μ) model
- Results & key findings

Motivating MET Data

Motivating MET data

The motivating MET data is from a series of 13 common bean *Phaselous vulgaris* variety trials conducted at 9 locations across Ethiopia in the 2022 and 2023 seasons

Image from Karyna Panchenko. CC0 1.0 license

Response variable is pod count per plant

Motivating MET Data

- Each trial contained 48 to 160 genotypes

Motivating MET Data

- Each trial contained 48 to 160 genotypes
- Each genotype was replicated 3 times using a row-column design using the `odw` R-package

Motivating MET Data

Genotype concurrence across trials

Trial	AN22	AN23	BK23	HU22	HW23	JM22	JM23	MK22	MK23	PW22	SK22	SK23	WK23	# of columns	# of rows
AN22	**160**													24	20
AN23	45	**135**												15	27
BK23	45	135	**135**											15	27
HU23	16	48	48	**48**										12	12
HW23	45	135	135	48	**135**									15	27
JM22	85	35	35	6	35	**85**								15	17
JM23	35	110	110	23	110	33	**110**							15	22
MK22	160	45	45	16	45	85	35	**160**						15	32
MK23	45	135	135	48	135	35	110	45	**135**					15	27
PW22	85	33	33	4	33	81	33	85	33	**85**				15	17
SK22	95	24	24	14	24	20	14	95	24	20	**95**			15	19
SK23	19	48	48	36	48	9	23	19	48	7	17	**48**		12	12
WK23	37	110	110	23	110	34	101	37	110	33	16	23	**110**	15	22

Mixed Models

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

Three strengths of mixed models (Smith *et al.* 2005)

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

Three strengths of mixed models (Smith *et al.* 2005)

- Ease with which incomplete data can be handled

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

Three strengths of mixed models (Smith *et al.* 2005)

- Ease with which incomplete data can be handled
- Ability to assume some sets of effects are random rather than fixed

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

Three strengths of mixed models (Smith *et al.* 2005)

- Ease with which incomplete data can be handled
- Ability to assume some sets of effects are random rather than fixed
- Ability to use more realistic variance models for random effects and residual error

The mixed model

The analysis of agricultural multi-environment trial (MET) data for continuous response variables is performed using linear mixed models (LMMs).

Three strengths of mixed models (Smith *et al.* 2005)

- Ease with which incomplete data can be handled
- Ability to assume some sets of effects are random rather than fixed
- Ability to use more realistic variance models for random effects and residual error

When the LMM assumptions are violated, an extension to a Generalised linear mixed model (GLMM) is required

Poisson distribution

Count data in agricultural research is **ubiquitous**

Poisson distribution

Count data in agricultural research is **ubiquitous**

- Most common distribution for count data is the Poisson(λ) distribution

Poisson distribution

Count data in agricultural research is **ubiquitous**

- Most common distribution for count data is the Poisson(λ) distribution
- Very restrictive assumption that $\mathrm{E}(y) = \mathrm{var}(y)$

Poisson distribution

Count data in agricultural research is **ubiquitous**

- Most common distribution for count data is the Poisson(λ) distribution
- Very restrictive assumption that $\mathrm{E}(y) = \mathrm{var}(y)$
- Some extensions to the Poisson distribution have been proposed
  - Generalised Poisson distribution
  - Tweedie distribution

Poisson distribution

Count data in agricultural research is **ubiquitous**

- Most common distribution for count data is the Poisson(λ) distribution
- Very restrictive assumption that $\mathrm{E}(y) = \mathrm{var}(y)$
- Some extensions to the Poisson distribution have been proposed
  - Generalised Poisson distribution
  - Tweedie distribution
- None of these proposed distributions can account for arbitrarily over and underdispersed count data

CMP$_\mu$ distribution

- The mean-parameterised Conway-Maxwell Poisson (CMP$_\mu$, Huang 2017) distribution overcomes these key issues

CMP$_\mu$ distribution

- The mean-parameterised Conway-Maxwell Poisson (CMP$_\mu$, Huang 2017) distribution overcomes these key issues
  1. CMP$_\mu$ distribution is part of the exponential family

CMP$_\mu$ distribution

- The mean-parameterised Conway-Maxwell Poisson (CMP$_\mu$, Huang 2017) distribution overcomes these key issues
  1. CMP$_\mu$ distribution is part of the exponential family
  2. The CMP$_\mu$ distribution can handle arbitrarily **over and underdispersed data** (Huang 2023)

CMP$_\mu$ distribution

- The mean-parameterised Conway-Maxwell Poisson (CMP$_\mu$, Huang 2017) distribution overcomes these key issues
  1. CMP$_\mu$ distribution is part of the exponential family
  2. The CMP$_\mu$ distribution can handle arbitrarily **over and underdispersed data** (Huang 2023)
  3. The dispersion parameter $\nu(\phi)$ is functionally independent of the location parameter μ (Huang & Rathouz 2017)

CMP$_\mu$ probability mass function

$$p(y;\mu,\nu) \propto \frac{\lambda(\mu,\nu)^y}{(y!)^\nu}, \quad y = 0,1,2,\ldots,$$

The mode parameter $\lambda(\mu,\nu)$ is obtained by solving

$$\sum_{y=0}^{\infty}(y-\mu)\frac{\lambda^y}{(y!)^\nu} = 0,$$

CMP$_\mu$ probability mass function

$$p(y;\mu,\nu) \propto \frac{\lambda(\mu,\nu)^y}{(y!)^\nu}, \quad y = 0,1,2,\ldots,$$

The mode parameter $\lambda(\mu,\nu)$ is obtained by solving

$$\sum_{y=0}^{\infty}(y-\mu)\frac{\lambda^y}{(y!)^\nu} = 0,$$

$$\phi \propto 1/\nu$$

Statistical Model

Statistical model

$$
\begin{aligned}
l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{y}|\boldsymbol{u}) &= \sum_{j=1}^{t} \sum_{i=1}^{m_j} \sum_{l=1}^{d_j} \log\big(p(y_{ijl}; \mu_{ijl}, \nu_j)\big) \\
\log(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_{\mathrm{g}}\boldsymbol{u}_{\mathrm{g}} + \boldsymbol{Z}_{\mathrm{o}}\boldsymbol{u}_{\mathrm{o}}, \\
\log(\boldsymbol{\nu}) &= \boldsymbol{X}\boldsymbol{\zeta},
\end{aligned}
$$

- $p(y_{ijl}; \mu_{ijl}, \nu_j)$ is the probability mass function for the CMP_{μ} distribution for the i^{th} genotype in the j^{th} environment (i.e. trial) within the l^{th} replicate block

Statistical model

$$
\begin{aligned}
l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{y}|\boldsymbol{u}) &= \sum_{j=1}^{t} \sum_{i=1}^{m_j} \sum_{l=1}^{d_j} \log\big(p(y_{ijl}; \mu_{ijl}, \nu_j)\big) \\
\log(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_{\mathrm{g}}\boldsymbol{u}_{\mathrm{g}} + \boldsymbol{Z}_{\mathrm{o}}\boldsymbol{u}_{\mathrm{o}}, \\
\log(\boldsymbol{\nu}) &= \boldsymbol{X}\boldsymbol{\zeta},
\end{aligned}
$$

- The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ are each of length t and denote the environment fixed effects for the mean and dispersion parameters respectively, each with corresponding design matrix $\boldsymbol{X}$
- t is the total number of environments

A factor analytic structure for the random G×E interaction effect

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\boldsymbol{u}_{\mathrm{g}}$ is the vector of random genotype by environment (G×E) interaction effects of length mt, and has a factor analytic structure of order k
- Denoted as an FA(k) model
- m is the total number of genotypes

A factor analytic structure for the random G×E interaction effect

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\boldsymbol{u}_{\mathrm{g}}$ is the vector of random genotype by environment (G×E) interaction effects of length mt, and has a factor analytic structure of order k
- Denoted as an FA(k) model
- m is the total number of genotypes
- $\boldsymbol{\Lambda}$ is a $t \times k$ matrix of environmental loadings

A factor analytic structure for the random G×E interaction effect

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\boldsymbol{u}_{\mathrm{g}}$ is the vector of random genotype by environment (G×E) interaction effects of length mt, and has a factor analytic structure of order k
- Denoted as an FA(k) model
- m is the total number of genotypes
- $\boldsymbol{\Lambda}$ is a $t \times k$ matrix of environmental loadings
- $\boldsymbol{f} \sim N(\boldsymbol{0}, \boldsymbol{I}_{mk})$ is a vector of genotype scores of length mk

A factor analytic structure for the random G×E interaction effect

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\boldsymbol{u}_{\mathrm{g}}$ is the vector of random genotype by environment (G×E) interaction effects of length mt, and has a factor analytic structure of order k
- Denoted as an FA(k) model
- m is the total number of genotypes
- $\boldsymbol{\Lambda}$ is a $t \times k$ matrix of environmental loadings
- $\boldsymbol{f} \sim N(\boldsymbol{0}, \boldsymbol{I}_{mk})$ is a vector of genotype scores of length mk
- $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \boldsymbol{\Psi} \otimes \boldsymbol{I}_m)$ is a vector of genetic regression residuals of length mt

A factor analytic structure for the random G×E interaction effects

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\mathrm{var}(\boldsymbol{u}_{\mathrm{g}}) = \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}\right) \otimes \boldsymbol{I}_m$

A factor analytic structure for the random G×E interaction effects

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\mathrm{var}(\boldsymbol{u}_{\mathrm{g}}) = \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}\right) \otimes \boldsymbol{I}_m$
- $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$ captures the 'repeatable' G×E interaction

A factor analytic structure for the random G×E interaction effects

$$\boldsymbol{u}_{\mathrm{g}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{f} + \boldsymbol{\delta}$$

- $\mathrm{var}(\boldsymbol{u}_{\mathrm{g}}) = \left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}\right) \otimes \boldsymbol{I}_m$
- $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$ captures the 'repeatable' G×E interaction
- $\boldsymbol{\Psi}$ is a $t \times t$ diagonal matrix containing the specific variances for each environment

Laplace Approximation estimation method

A difficulty arising with GLMMs is that inference from the marginal log-likelihood is analytically intractable

Laplace Approximation estimation method

A difficulty arising with GLMMs is that inference from the marginal log-likelihood is analytically intractable

The Laplace Approximation estimation method acquires a good **'balance'** between estimation accuracy and computational speed (Pinheiro *et al.* 1995).

Laplace Approximation estimation method

A difficulty arising with GLMMs is that inference from the marginal log-likelihood is analytically intractable

The Laplace Approximation estimation method acquires a good **'balance'** between estimation accuracy and computational speed (Pinheiro *et al.* 1995).

- Computationally infeasible for complex variance structures

Laplace Approximation estimation method

A difficulty arising with GLMMs is that inference from the marginal log-likelihood is analytically intractable

The Laplace Approximation estimation method acquires a good **'balance'** between estimation accuracy and computational speed (Pinheiro *et al.* 1995).

- Computationally infeasible for complex variance structures
- This has been resolved recently with the `glmmTMB` `R`-package (Brooks *et al.* 2017)
- `glmmTMB` uses **automatic differentiation** to speed up the computation of high dimensional gradient functions (Griewank and Walther, 2008)

REML correction for GLMM

The `glmmTMB` R-package has the option to use the LA with a REML adjustment.

REML correction for GLMM

The `glmmTMB` R-package has the option to use the LA with a REML adjustment.

- The REML adjustment is obtained by marginalising with respect to $\boldsymbol{\beta}$ as well as $\boldsymbol{u}$ (Maestrini *et al.* 2024)

$$l_{\text{REML}}(\gamma, \phi; \boldsymbol{y}) \approx l(\hat{\boldsymbol{\beta}}, \gamma, \phi, \tilde{\boldsymbol{u}}; \boldsymbol{y}) + (n - t)\log\sqrt{2\pi} - \frac{1}{2}|\boldsymbol{H}^*|$$

- $\boldsymbol{H}^*$ is the Hessian of $l(\hat{\boldsymbol{\beta}}, \gamma, \phi, \tilde{\boldsymbol{u}}; \boldsymbol{y})$

Aim

The proposed statistical model for the analysis of MET count data is referred to as the FA-CMP$_\mu$ model

- Model G$\times$E interaction effects using FA variance structure

Aim

The proposed statistical model for the analysis of MET count data is referred to as the FA-CMP$_\mu$ model

- Model G$\times$E interaction effects using FA variance structure
- Model heterogeneous dispersion using the CMP$_\mu$ distribution

Aim

The proposed statistical model for the analysis of MET count data is referred to as the FA-CMP$_\mu$ model

- Model G$\times$E interaction effects using FA variance structure
- Model heterogeneous dispersion using the CMP$_\mu$ distribution
- Enables partitioning of genetic sources of variation from non-genetic sources and dispersion

Aim

The proposed statistical model for the analysis of MET count data is referred to as the FA-CMP$_\mu$ model

- Model G$\times$E interaction effects using FA variance structure
- Model heterogeneous dispersion using the CMP$_\mu$ distribution
- Enables partitioning of genetic sources of variation from non-genetic sources and dispersion
- Goal is genotypic selection by maximising prediction accuracy of $\boldsymbol{u}_{\text{g}}$

Model implementation

FA models of various order were fit to model the G$\times$E interaction effects

- Fitted initial model using a diagonal variance structure before fitting the FA(1) model

Model implementation

FA models of various order were fit to model the G$\times$E interaction effects

- Fitted initial model using a diagonal variance structure before fitting the FA(1) model
- Initial values from the preceding model were used as starting values

Model implementation

FA models of various order were fit to model the G$\times$E interaction effects

- Fitted initial model using a diagonal variance structure before fitting the FA(1) model
- Initial values from the preceding model were used as starting values
- Random row and column terms fitted for each trial to adjust for spatial field trend through $\boldsymbol{u}_{\text{o}}$

Results & key findings

Results of MET analysis

FA model of order 3 was the best fit as per the AIC criteria

- FA(3) model explains 75.1% of the total G×E interaction effects.

$$\text{AIC} = -2l_{\text{LA, REML}}(\boldsymbol{\gamma}, \boldsymbol{\phi}; \boldsymbol{y}) + 2q$$

G×E variance structure	Number of non-boundary parameters (q)	AIC
Diagonal	49	25857
FA($k = 1$)	62	25695
FA($k = 2$)	74	25694
FA($k = 3$)	79	**25688**
FA($k = 4$)	89	25695
FA($k = 5$)	94	25701
Unstructured	127	25761

Results of MET analysis

DELIVERING FOR QUEENSLAND | Queensland Government

Trial	Trial log Mean $\hat{\beta}_j$	Trial Dispersion $1/\hat{\nu}_j$	Sqrt Genetic Variance $\hat{\sigma}_{g_j}$
AN22	3.29	1.88	0.18
AN23	2.98	1.62	0.13
BK23	2.32	**0.72**	**0.19**
HU23	3.00	**0.18**	0.08
HW23	2.74	0.72	0.10
JM22	2.93	0.69	0.10
JM23	2.63	1.32	0.17
MK22	3.01	1.43	0.17
MK23	3.09	**2.51**	**0.22**
PW22	2.19	0.39	0.18
SK22	2.75	0.38	0.12
SK23	**3.33**	**0.61**	0.08
WK23	2.86	**0.82**	**0.20**

- All trials have genetic variance

Results of MET analysis

DELIVERING FOR QUEENSLAND | Queensland Government

Trial	Trial log Mean $\hat{\beta}_j$	Trial Dispersion $1/\hat{\nu}_j$	Sqrt Genetic Variance $\hat{\sigma}_{g_j}$
AN22	3.29	1.88	0.18
AN23	2.98	1.62	0.13
BK23	2.32	**0.72**	**0.19**
HU23	3.00	**0.18**	0.08
HW23	2.74	0.72	0.10
JM22	2.93	0.69	0.10
JM23	2.63	1.32	0.17
MK22	3.01	1.43	0.17
MK23	3.09	**2.51**	**0.22**
PW22	2.19	0.39	0.18
SK22	2.75	0.38	0.12
SK23	**3.33**	**0.61**	0.08
WK23	2.86	**0.82**	**0.20**

- All trials have genetic variance
- Some trials have strong under dispersion

Results of MET analysis

DELIVERING FOR QUEENSLAND | Queensland Government

Trial	Trial log Mean $\hat{\beta}_j$	Trial Dispersion $1/\hat{\nu}_j$	Sqrt Genetic Variance $\hat{\sigma}_{g_j}$
AN22	3.29	1.88	0.18
AN23	2.98	1.62	0.13
BK23	2.32	**0.72**	**0.19**
HU23	3.00	**0.18**	0.08
HW23	2.74	0.72	0.10
JM22	2.93	0.69	0.10
JM23	2.63	1.32	0.17
MK22	3.01	1.43	0.17
MK23	3.09	**2.51**	**0.22**
PW22	2.19	0.39	0.18
SK22	2.75	0.38	0.12
SK23	**3.33**	**0.61**	0.08
WK23	2.86	**0.82**	**0.20**

- All trials have genetic variance
- Some trials have strong under dispersion
- Trial with highest mean count is underdispersed

Results of MET analysis

DELIVERING FOR QUEENSLAND | Queensland Government

Trial	Trial log Mean $\hat{\beta}_j$	Trial Dispersion $1/\hat{\nu}_j$	Sqrt Genetic Variance $\hat{\sigma}_{g_j}$
AN22	3.29	1.88	0.18
AN23	2.98	1.62	0.13
BK23	2.32	**0.72**	**0.19**
HU23	3.00	**0.18**	0.08
HW23	2.74	0.72	0.10
JM22	2.93	0.69	0.10
JM23	2.63	1.32	0.17
MK22	3.01	1.43	0.17
MK23	3.09	**2.51**	**0.22**
PW22	2.19	0.39	0.18
SK22	2.75	0.38	0.12
SK23	**3.33**	**0.61**	0.08
WK23	2.86	**0.82**	**0.20**

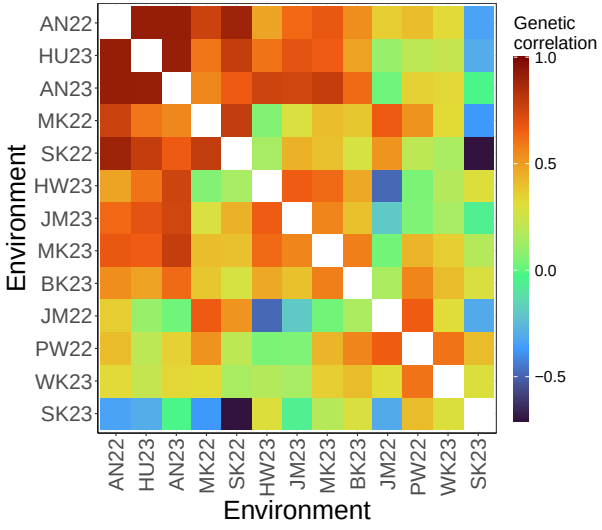
- All trials have genetic variance
- Some trials have strong under dispersion
- Trial with highest mean count is underdispersed
- Trial with highest dispersion also has the highest genetic variance

Results of MET analysis

Trial	Trial log Mean $\hat{\beta}_j$	Trial Dispersion $1/\hat{\nu}_j$	Sqrt Genetic Variance $\hat{\sigma}_{g_j}$
AN22	3.29	1.88	0.18
AN23	2.98	1.62	0.13
BK23	2.32	**0.72**	**0.19**
HU23	3.00	**0.18**	0.08
HW23	2.74	0.72	0.10
JM22	2.93	0.69	0.10
JM23	2.63	1.32	0.17
MK22	3.01	1.43	0.17
MK23	3.09	**2.51**	**0.22**
PW22	2.19	0.39	0.18
SK22	2.75	0.38	0.12
SK23	**3.33**	**0.61**	0.08
WK23	2.86	**0.82**	**0.20**

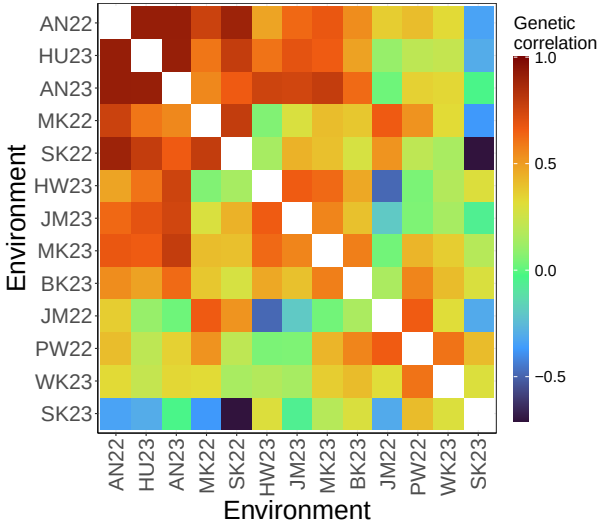
- All trials have genetic variance
- Some trials have strong under dispersion
- Trial with highest mean count is underdispersed
- Trial with highest dispersion also has the highest genetic variance
- Trials with next two highest genetic variances are underdispersed

Results of MET analysis - Heatmap



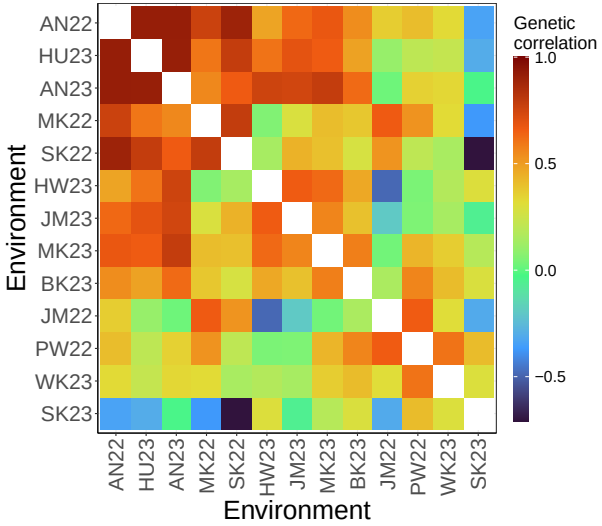
- Genetic correlations obtained from covariance matrix G_e
- Trials ordered by clustering algorithm

Results of MET analysis - Heatmap



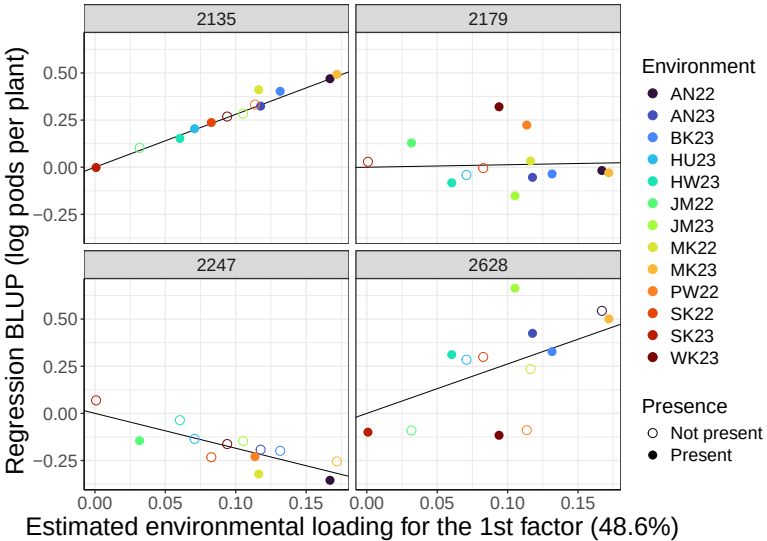
- Genetic correlations obtained from covariance matrix G_e
- Trials ordered by clustering algorithm
- Cluster of trials with high genetic correlation on the top left

Results of MET analysis - Heatmap



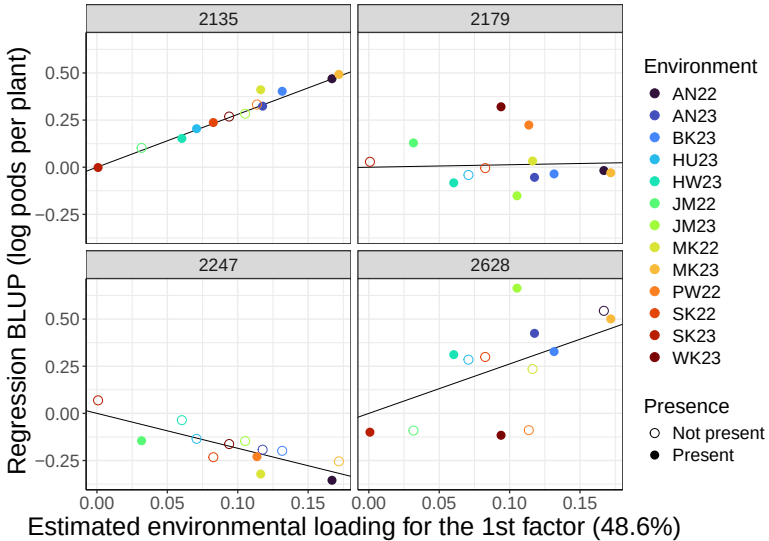
- Genetic correlations obtained from covariance matrix G_e
- Trials ordered by clustering algorithm
- Cluster of trials with high genetic correlation on the top left
- Trials on the bottom right generally have low genetic correlations

Latent regression plot



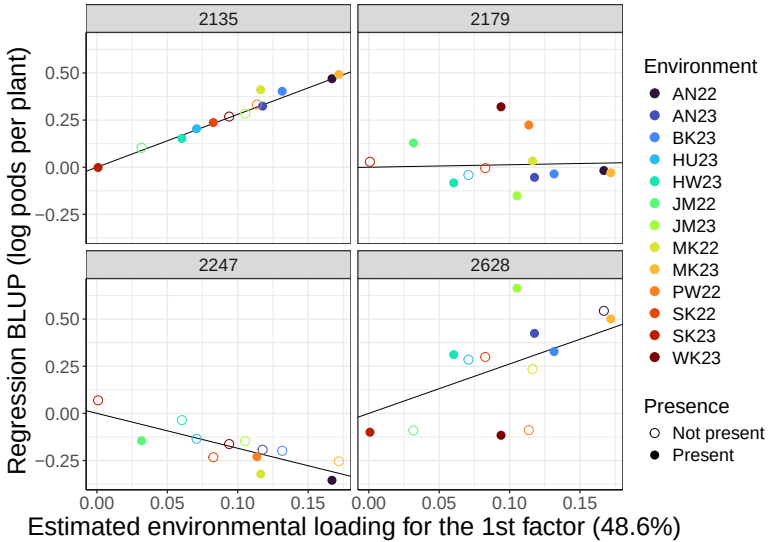
- Subset of four genotypes

Latent regression plot



- Subset of four genotypes
- All rotated loadings > 0

Latent regression plot



- Subset of four genotypes
- All rotated loadings > 0
- Some genotypes have higher 'lack of fit'

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean
- Use an FA structure to model the G$\times$E interaction effects

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean
- Use an FA structure to model the G$\times$E interaction effects
- Uses the `glmmTMB` R-package to fit the model

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean
- Use an FA structure to model the G$\times$E interaction effects
- Uses the `glmmTMB` R-package to fit the model
  - Automatic differentiation to make model fitting computationally feasible

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean
- Use an FA structure to model the G$\times$E interaction effects
- Uses the `glmmTMB` R-package to fit the model
  - Automatic differentiation to make model fitting computationally feasible
  - Laplace approximation estimation method to reduce bias

Summary

The FA-CMP$_\mu$ model is proposed to analyse MET count data using a GLMM framework

- Assumes the data follows a CMP$_\mu$ distribution
  - Enables partitioning of dispersion from genetic and non-genetic variation
  - Trial dispersion independent of trial mean
- Use an FA structure to model the G$\times$E interaction effects
- Uses the `glmmTMB` R-package to fit the model
  - Automatic differentiation to make model fitting computationally feasible
  - Laplace approximation estimation method to reduce bias
  - REML correction to further reduce estimation bias

Acknowledgements

- Dr. Alan Huang and Dr. Alison Kelly for their ongoing support, expertise, & mentorship
- Queensland Department of Primary Industries for financial support & time to work on the project

Acknowledgements

- The National Lowland Pulse Research Program at the Ethiopian Institute for Agricultural Research (EIAR) for conducting the trials & offering essential facilities, data & support for this study.
- The MERCI project team at both EIAR and the University of Queensland within QAAFI, funded by the Gates Foundation

References

- Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... and Bolker, B. M. (2017). `glmmTMB` balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling.
- Griewank, A. and Walther, A. (2008). Evaluating derivatives: principles and techniques of algorithmic differentiation. *SIAM*.
- Huang, A. (2017). Mean-parameterized Conway-Maxwell-Poisson regression models for dispersed counts. *Statistical Modelling*, 17, 359-380.
- Huang, A. (2023). On arbitrarily underdispersed discrete distributions. *The American Statistician*, 77, 29-34.
- Huang, A. and Rathouz, P.J. (2017). Orthorogonality of the mean and error distribution in generalized linear models. *Communications in Statistics-Theory and Methods*, 46, 3290-3296.
- Maestrini, L., Hui, F. K., and Welsh, A. H. (2024). Restricted maximum likelihood estimation in generalized linear mixed models. *arXiv preprint arXiv:2402.12719*
- Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4, 12-35.
- Smith, A., Cullis, B. R., and Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agriculltural Science*, 143, 449-462.