

Mixed models for GxE data (Part II): Modelling QTL x E effects

Marcos Malosetti & Piter Bijma
Wageningen University and Research
Wageningen, October 2015



Contents

- QTLxE models by adding genetic predictors in the LMM
- QTL mapping: model building (a strategy)
- QTLxE mapping examples



Modelling mean and VCOV for MET data in LMM

$$\underline{y}_{ij} = \mu_{ij} + \underline{\epsilon}_{ij} \quad \underline{\epsilon}_{ij} \sim MVN(0, \Sigma)$$

□ Aim of statistical modelling for MET/GxE data

- μ_{ij} : **predictable**
 - **Goal:** Describe μ_{ij} as much as possible in terms of single indexed parameters
 - = Separate G_i and E_j terms
- $VCOV(\underline{\epsilon}_{ij}) = \Sigma$: **unpredictable**

Modelling QTL and QTLxG

□ Molecular markers can be fitted as covariables:

- $x_i = \{0, 1, 2\}$ or $x_i = \{-1, 0, 1\}$
- $MM \rightarrow x_i = 1$
- $Mm \rightarrow x_i = 0$
- $mm \rightarrow x_i = -1$

□ We want to test whether the variation at molecular markers is related to phenotypic variation

- In the main effect (G_i)
- In the specific environments (GE_{ij})

Modelling GxE in terms of QTLs

$$\begin{aligned} \underline{y}_{ij} &= \mu + E_j + (\underline{G}_i + \underline{GE}_{ij}) \\ \underline{y}_{ij} &= \mu + E_j + x_i \alpha + \underline{G}_i^* + x_i \alpha_j^* + \underline{GE}_{ij}^* \\ \underline{y}_{ij} &= \mu + E_j + x_i \alpha_j + \underline{GE}_{ij}^* \end{aligned}$$

- α : QTL main effect
- α_j^* : QTLxE = deviation from QTL main effect
- α_j : Environment specific QTL effect
- Focus on environment-specific QTL effects α_j and the modelling of the random \underline{GE}_{ij}^*
- $\underline{GE}_{ij}^* \sim \text{MVN}(0, \Sigma_{GE})$; $\Sigma_{GE} = \Sigma_G \otimes \Sigma_E$

QTL mapping; a strategy

1. Find a VCOV model for the GGE (=G+GE) variation in the phenotypic data
2. Test for QTLs by single-SNP models
3. Perform backward selection on the full set of signific. SNP
4. Estimate effects in final multi-QTL model
5. Option: Identify environmental covariables related to environment-specific QTL effects

QTL mapping: model building

1. $\underline{y}_{ij} = \mu_i + E_j + \underline{\epsilon}_{ij}$ Assess / find a model for $\underline{\epsilon}_{ij}$
2. $\underline{y}_{ij} = \mu_i + E_j + x_i \alpha_j + \underline{\epsilon}_{ij}$ Genome-wide QTL search
(Single SNP)
3. $\underline{y}_{ij} = \mu_i + E_j + \sum x_i \alpha_j + \underline{\epsilon}_{ij}$ Backwards elimination
4. $\underline{y}_{ij} = \mu_i + E_j + \sum_{\text{significant}} x_i \alpha_j + \underline{\epsilon}_{ij}$ Final multi-QTL model

Testing: main effects and GxE

$$\underline{y}_{ij} = \mu_i + E_j + \sum x_i \alpha_j + \underline{\epsilon}_{ij}$$

- For each QTL a specific test can be made to assess whether QTLxG is present or not:

$$\underline{y}_{ij} = \mu_i + E_j + x_i \alpha_j + \underline{\epsilon}_{ij} \quad \text{QQE}$$

$$\underline{y}_{ij} = \mu_i + E_j + x_i \alpha + x_i \alpha_j^* + \underline{\epsilon}_{ij} \quad \text{Q+QxE}$$

$$\underline{y}_{ij} = \mu_i + E_j + x_i \alpha + \underline{\epsilon}_{ij} \quad \text{Q}$$

QTLxE: CIMMYT drought stress in maize

- Response: Yield
- Environments
 - 8 trials = 8 managed stress environments, intermediate and severe drought stress (IS, SS), low and high nitrogen (LN, HN), no stress
 - 1992, 1994, 1996
 - 2 locations (TI, PR)
 - Winter and summer seasons
- Genotypes
 - 211 F2 derived F3 lines
- Covariables
- Genotypes
 - 132 marker loci
- Min. and max. temperature, radiation, rain and number of sun hours for vegetative, flowering and grain filling stages



Maize data: phenotypic data and genetic predictors

GI	y_IS94a	y_SS94a	y_HN96b	y_LN96b	y_LN98a	y_IS92a	y_NS92a	y_SS92a
1	337	448	657	71	145	672	1260	493
2	603	332	407	140	88	732	1143	438
3	342	267	574	108	171	680	1152	409
4	208	22	343	0	110	554	767	369
5	453	26	496	03	166	754	895	427
6	322	40	336	36	251	594	1144	267
7	336	32	578	47	350	678	1094	376
8	144	4	531	57	219	384	710	158
9	264	30	601	124	194	464	1327	202
10	506	557	669	147	150	611	1113	495
11	361	302	353	63	180	539	993	349
12	285	242	254	14	211	626	1024	356
13	594	591	604	187	22	705	901	413
14	662	398	708	254	282	501	1082	278
15	229	358	346	39	85	614	844	316

GI	mk_add[1]	mk_add[2]	mk_add[3]	mk_add[4]	mk_add[5]	mk_add[6]	mk_add[7]	mk_add[8]	mk_add[9]	mk_add[10]	mk_add[11]	mk_add[12]	mk_add[13]	mk_add[14]	mk_add[15]	mk_add[16]	mk_add[17]	mk_add[18]	mk_add[19]
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
9	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Construction of predictors for additive genetic effects *at* marker positions:

$$x_i(\text{MM}) = 1; x_i(\text{Mm}) = 0; x_i(\text{mm}) = -1$$



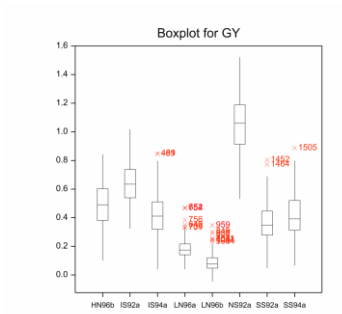
Summary statistics per environment

Summary statistics for GY: Trial HN96b

Number of observations = 211
 Number of missing values = 0
 Mean = 0.485
 Median = 0.43
 Minimum = 0.101
 Maximum = 0.843
 Lower quartile = 0.381
 Upper quartile = 0.602
 Standard deviation = 0.150
 Variance = 0.0225

Summary statistics for GY: Trial IS92a

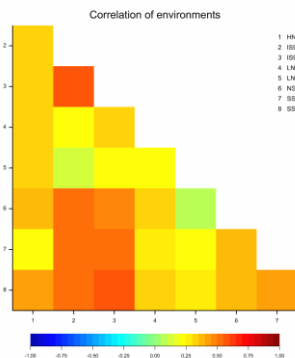
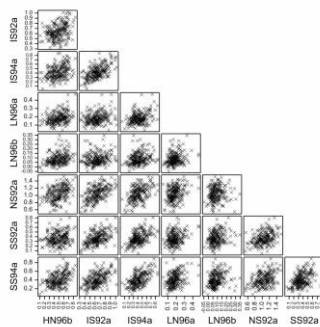
Number of observations = 211
 Number of missing values = 0
 Mean = 0.640
 Median = 0.634
 Minimum = 0.325
 Maximum = 1.018
 Lower quartile = 0.539
 Upper quartile = 0.737
 Standard deviation = 0.146
 Variance = 0.0214



- Different mean performance
- Heterogeneity of variation



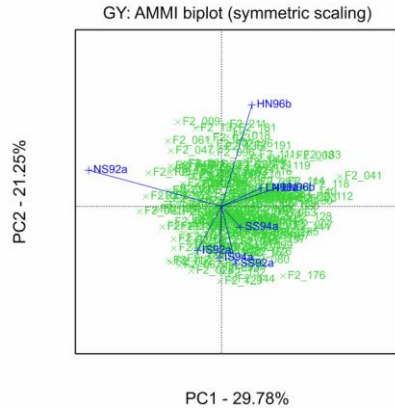
Correlations between environments



- Heterogeneity of correlation (=GxE)



AMMI biplot



□ NS92a and HN96b large contribution to GxE



Selection of VCOV matrix

Summary

Trait: GY

Model	AIC	SIC	Deviance	NParameters
FA	-5739	-5686	-5771	16
FA2	-5755	-5678	-5801	23
OUTSIDE	-5687	-5656	-5705	9
UNSTRUCTURED	-5753	-5633	-5825	36
HCS	-5518	-5488	-5536	9
CS	-5293	-5286	-5297	2
DIAGONAL	-5304	-5277	-5320	8
IDENTITY	-4923	-4920	-4925	1

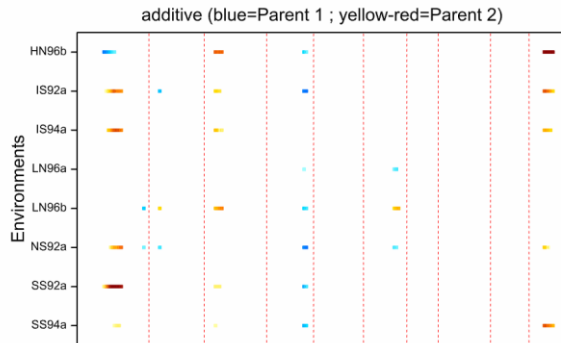
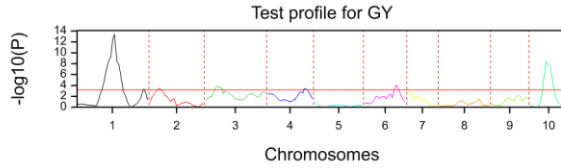
Best model: FA (on basis of criterion SIC)



QTL search: SIM (single SNP interval mapping)

$$\underline{y}_{ij} = \mu_i + E_j + x_i \alpha_j + \epsilon_{ij}$$

- QTL on chr 1 large GxE (cross-over).
- QTL on chr 10 large GxE (but no cross-over).
- QTL on chr 4 consistent effect (no QTLxE).



Selection final multi-QTL model

Summary

Trait: GY
 Population type: F2
 Number of genotypes: 211
 Number of environments: 8
 Number of linkage groups: 10
 Number of markers: 122
 Variance-covariance model: FA

$$\underline{y}_{ij} = \mu_i + E_j + \sum_{q \in Q} x_{iq} \alpha_{jq} + \epsilon_{ij}$$

List of QTLs

Locus no.	Locus name	Linkage group	Position	$-\log_{10}(P)$	QTLxE
19	L085	1	141.00	13.767	yes
42	C2P36	2	35.90	4.670	yes
73	L035	3	55.70	4.663	yes
112	L071	4	136.60	3.527	no
159	L043	6	125.00	3.711	yes
237	C10P60	10	60.15	8.318	yes

- 6 QTL in the final model
- All show significant QTLxE, except on chr 4.



QTL on chromosome 1

QTL (Locus name): L085

Location: linkage group 1 position 141

Environment	Effect	High value allele	s.e.	P	%Expl. var.	CI_LL	CI_UL
HN96b	0.039	P1	0.013	0.003	3.3	124.21	157.79
IS92a	0.055	P2	0.013	0.000	7.0	124.21	157.79
IS94a	0.056	P2	0.014	0.000	7.1	124.21	157.79
LN96a	0.000	P2	0.007	0.983	0.0	*	*
LN96b	0.001	P2	0.006	0.849	0.0	*	*
NS92a	0.064	P2	0.019	0.001	5.2	124.21	157.79
SS92a	0.072	P2	0.012	0.000	14.9	124.21	157.79
SS94a	0.027	P2	0.015	0.063	1.6	*	*

□ Strong QTLxE

- Cross-over type of interaction (HN96b vs rest)
- No effect in LN96a/b



QTL on chromosome 10

QTL (Locus name): C10P60

Location: linkage group 10 position 60.15

Environment	Effect	High value allele	s.e.	P	%Expl. var.	CI_LL	CI_UL
HN96b	0.093	P2	0.014	0.000	19.2	48.26	72.04
IS92a	0.056	P2	0.014	0.000	7.3	48.26	72.04
IS94a	0.042	P2	0.015	0.004	4.1	48.26	72.04
LN96a	0.010	P2	0.007	0.160	1.1	*	*
LN96b	0.008	P2	0.006	0.186	0.9	*	*
NS92a	0.047	P2	0.020	0.018	2.8	48.26	72.04
SS92a	0.013	P2	0.013	0.287	0.5	*	*
SS94a	0.060	P2	0.015	0.000	8.0	48.26	72.04

□ Also QTLxE, but no crossover



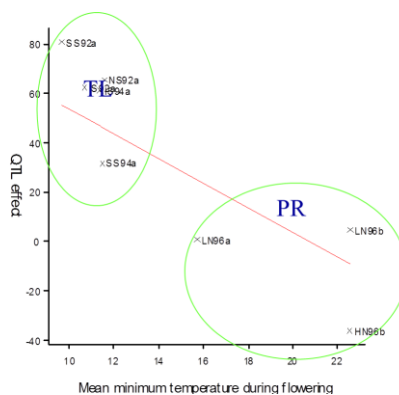
QTL effects in relation to env covariables

$$\underline{y}_{ij} = \mu + E_j + x_i \alpha_j + \underline{GE}_{ij}^*$$

$$\underline{y}_{ij} = \mu + E_j + x_i (\alpha_q + \beta_q z_j + \underline{a}_j^*) + \underline{GE}_{ij}^*$$

- Decompose QTL effect in:
 - An intercept across all environments (α_q)
 - Part associated with environmental covariable (z_j)
 - Reaction norm with slope β_q .
 - Residual QTL effect (random \underline{a}_j^*).
- Prediction possible within the range of values of z_j

Regression of QTLx E on min. temperature during flowering



- QTL on chromosome 1 reacts to temperature.
 - Now we “understand” the Gx E -mechanism

Conclusions

- ❑ QTLx E -mapping: find the genes causing the Gx E -interaction
- ❑ Regression of QTL-effect on explicit environmental covariables
 - “Understand” the Gx E -mechanism
 - Prediction of performance in other environments as a function of QTL-genotype
 - Use: Preselection of high-potentials before field testing
- ❑ Should animal breeders be doing this kind of thing?