Improving genomic-enabled prediction accuracy by modeling the genotype-by-environment interaction for quality traits in Kansas wheat

Reka Howard, Diego Jarquin, Jose Crossa,
Sarah Battenfield, Jesse Poland, Carlos Guzman

Department of Statistics
University of Nebraska – Lincoln, USA

Outline

- **Introduction**
  - Kansas wheat data
- **GP Model Development**
  - The nature of the problem
  - Assumptions for main and interaction effects
- **Empirical Assessment**
  - Cross-Validation Schemes
  - Assessment of predictive ability
- **Results**
  - **CV2**
  - **CV1**
  - **CV0**
  - **CV00**

Experimental Description

- 415 unique lines
- 666 lines
- 2006 - 2014
- 7 locations
- 17 environments (year x location)
- Lines were genotyped at 4794 SNPs

Experimental Description

Year	Location	Environment	Sample Size
2006	Central Kansas	2006_CKS	34
2007	Central Kansas	2007_CKS	8
2008	Central Kansas	2008_CKS	68
2009	Central Kansas	2009_CKS	102
2010	Central Kansas	2010_CKS	43
2011	Central Kansas	2011_CKS	35
2012	Belleville	2012_BE	65
2012	Gypsum	2012_GY	6
2012	Kansas	2012_K	6
2012	Manhattan	2012_M	6
2012	McPherson	2012_MP	65
2013	Gypsum	2013_GY	45
2013	Kansas	2013_K	1
2013	McPherson	2013_MP	40
2014	Kansas	2014_CKS	133
2014	Manhattan	2014_M	2
2014	Sumner	2014_SU	7

Phenotypic Data

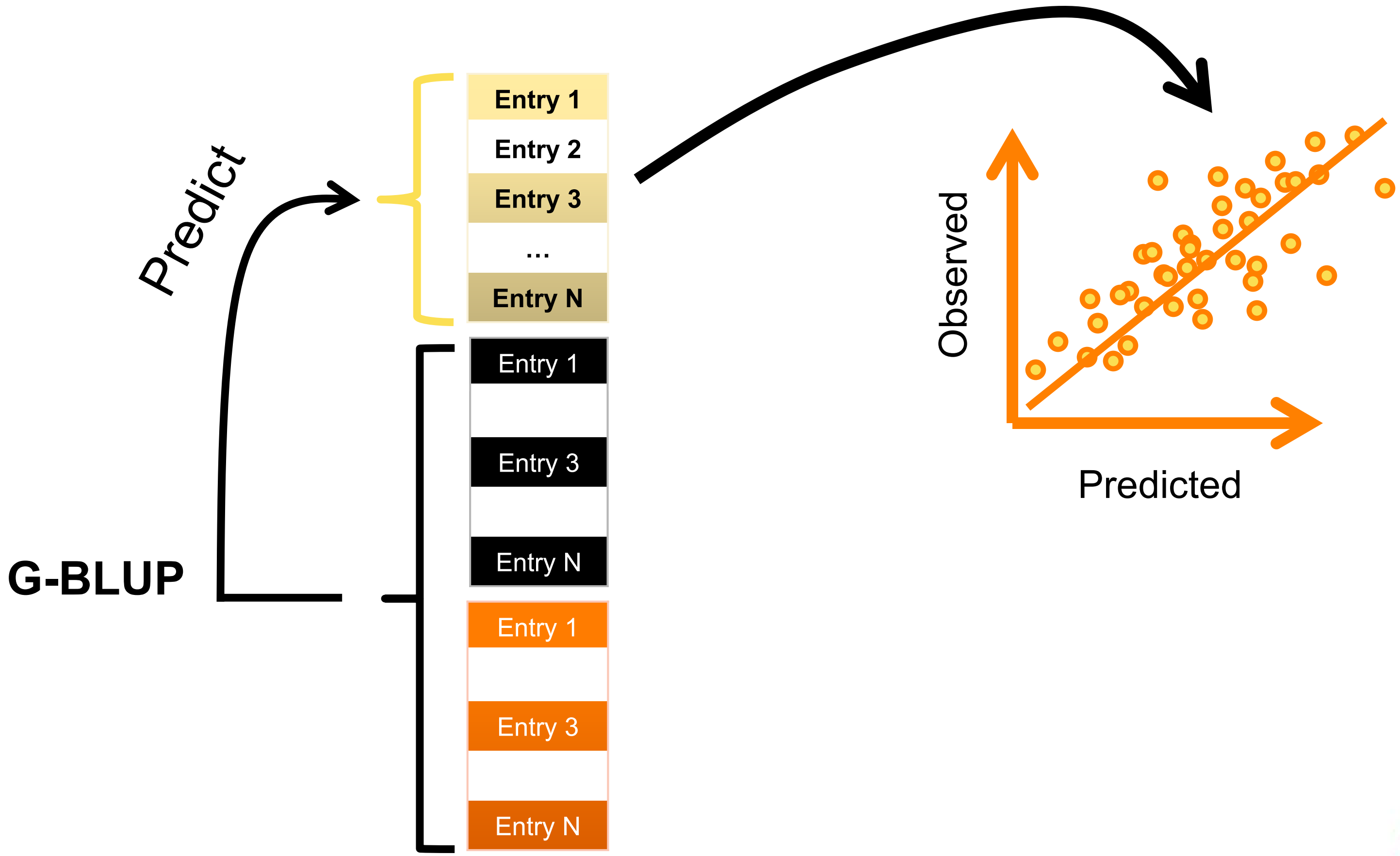
- **Evaluation across all environments included:**
- TESTWT
- AVGWT
- AVGDIA
- GRNPRO
- FLRPRO
- MIXABS
- MIXTIM
- BAKEABS
- BAKETIM
- LOFVOL
- IBAKETIM
- IMIXTIM

Prediction of Complex Traits: The Nature of the Problem

- Many important traits in plants are complex
- They are affected by large numbers of small-effect genes and large number of environmental factors
- Genes and environmental factors may interact in complex ways
- Challenges:
  - How to cope with highly dimensional nature of genomic and environmental information?
  - How models can capture complex forms of interactions among large number of factors?

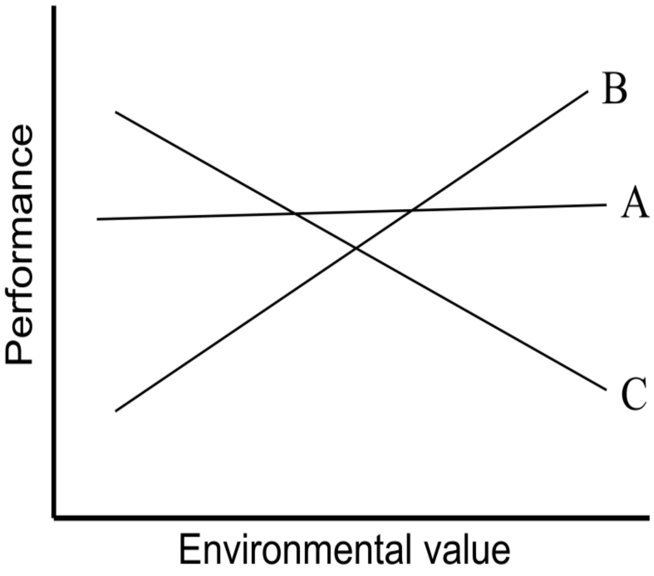
Methods (GS in a nutshell)

- In principle, use of phenotypes and genotypes in a training set to calibrate models and perform predictions in unobserved genotypes.

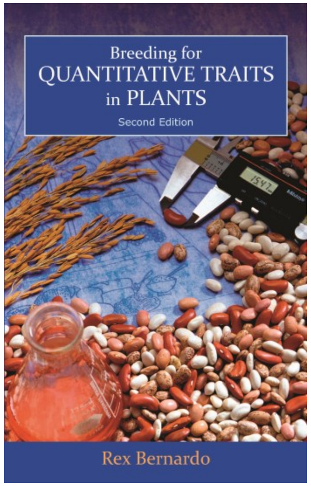


Introduction: The nature of the problem predicting complex traits in presence of G × E

- Options for dealing with GxE



1. Ignore it
2. Reduce it
3. Exploit it



Model Development

Accommodating Main Effects of SNPs, Environmental Covariates (ECs) and interactions between SNPs and ECs (Jarquin et al. 2014):

$$y_i = \mu + w_i + g_i + gw_i + \varepsilon_i$$

$$\varepsilon_i \overset{IID}{\sim} \mathrm{N}\left(0, \sigma_\varepsilon^2\right)$$

$$\mathbf{g} = \{g_i\} = \mathbf{X\beta} \qquad \mathbf{g} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{G}\sigma_g^2\right) \qquad \mathbf{G} = \frac{1}{p}\mathbf{XX}'$$

GRM or Kinship matrix

$$\mathbf{w} = \{w_i\} \qquad \mathbf{w} \sim \mathrm{N}\left(0, \mathbf{\Omega}\sigma_W^2\right) \qquad \mathbf{\Omega} = \frac{1}{Q}\mathbf{WW}'$$

ERM accounting for environmental factors

$$\mathbf{wg} = \{wg_i\} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{G}\#\mathbf{\Omega}\cdot\sigma_{GW}^2\right)$$

IRM accounting for GxE

Empirical Assessment

Empirical Assessment – CV2:

- **Predicting Performance of Lines Captured in Other Environments**

	E1	E2	E3	E4	E5
Line 1	Y_{11}	NA	Y_{13}	Y_{14}	Y_{15}
Line 2	Y_{21}	Y_{22}	NA	Y_{24}	Y_{25}
Line 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	NA
Line 4	Y_{41}	Y_{42}	Y_{43}	NA	Y_{45}
Line 5	NA	Y_{52}	Y_{53}	Y_{54}	Y_{55}

- **We used a 5 fold design**
- **50 Reps**

Empirical Assessment - CV1:

- Predicting performance of new developed lines through relationships with others

	E1	E2	E3	E4	E5
Line 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}
Line 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}
Line 3	NA	NA	NA	NA	NA
Line 4	Y_{41}	Y_{42}	Y_{43}	Y_{44}	Y_{45}
Line 5	Y_{51}	Y_{52}	Y_{53}	Y_{54}	Y_{55}

- We used a 5 fold design
- 50 Reps

Empirical Assessment – CV0:

- **Predicting performance of unobserved environments**

	E1	E2	E3	E4	E5
Line 1	Y_{11}	Y_{12}	NA	Y_{14}	Y_{15}
Line 2	Y_{21}	Y_{22}	NA	Y_{24}	Y_{25}
Line 3	Y_{31}	Y_{32}	NA	Y_{34}	Y_{35}
Line 4	Y_{41}	Y_{42}	NA	Y_{44}	Y_{45}
Line 5	Y_{51}	Y_{52}	NA	Y_{54}	Y_{55}

- **Leaving one environment out at the time**

Empirical Assessment – CV00:

- **Predicting performance of unobserved environments**

	E1	E2	E3	E4	E5
Line 1	Y_{11}	Y_{12}	NA	Y_{14}	Y_{15}
Line 2	Y_{21}	Y_{22}	NA	Y_{24}	Y_{25}
Line 3	NA	NA	NA	NA	NA
Line 4	Y_{41}	Y_{42}	NA	Y_{44}	Y_{45}
Line 5	Y_{51}	Y_{52}	NA	Y_{54}	Y_{55}

- **Leaving one environment out at the time**

CV2: Incomplete field trials (50 reps)

LOFVOL

Environment	Sample Size	CV2		
		L+E	L+E+G	L+E+G+GE
2009_CKS	102	-0.211	-0.158	-0.111
2010_CKS	43	0.092	0.069	0.373
2011_CKS	35	0.166	0.311	0.514
2012_BE	65	0.411	0.563	0.646
2012_GY	6	0.712	0.740	0.672
2012_K	6	-0.257	-0.216	-0.445
2012_M	6	-0.093	-0.075	-0.425
2012_MP	66	0.348	0.379	0.413
2008_CKS	68	0.048	0.142	0.283
2006_CKS	34	0.327	0.307	0.283
2013_GY	46	0.411	0.430	0.368
2013_K	5	-0.190	-0.319	-0.504
2013_MP	40	0.138	0.130	0.146
2007_CKS	8	-0.739	-0.725	-0.759
2014_CKS	133	-0.149	0.121	0.176
2014_SU	7	-0.558	-0.582	-0.434
2014_M	2	-1.000	-1.000	-1.000
2005_CKS	9	0.511	0.603	0.634

CV2: Incomplete field trials (50 reps)

	CV2		
	E+L	E+GL+G	E+L+G+GE
AVGDIA	0.241	0.328	0.328
AVGWT	0.142	0.269	0.274
BAKEABS	-0.011	0.030	0.058
BAKETIM	0.482	0.521	0.529
FLRPRO	0.012	0.115	0.151
GRNPRO	0.108	0.251	0.267
IBAKETIM	0.506	0.533	0.536
IMIXTIM	0.452	0.484	0.500
LOFVOL	0.075	0.169	0.232
MIXABS	-0.043	0.010	0.036
MIXTIM	0.462	0.502	0.514
TESTWT	0.180	0.296	0.316

CV1: Newly developed lines (50 reps)

LOFVOL

Environment	Sample Size	CV1		
		L+E	L+E+G	L+E+G+GE
2009_CKS	102	-0.158	-0.049	0.024
2010_CKS	43	-0.283	-0.074	0.186
2011_CKS	35	-0.410	0.292	0.352
2012_BE	65	-0.228	0.354	0.495
2012_GY	6	-0.586	0.654	-0.172
2012_K	6	-0.185	0.277	0.231
2012_M	6	0.056	0.074	0.074
2012_MP	66	-0.197	0.212	0.257
2008_CKS	68	-0.262	0.066	0.228
2006_CKS	34	-0.340	0.076	-0.063
2013_GY	46	-0.257	0.178	0.156
2013_K	5	-0.567	0.089	-0.393
2013_MP	40	-0.272	-0.176	-0.202
2007_CKS	8	-0.595	-0.686	-0.705
2014_CKS	133	-0.160	0.110	0.164
2014_SU	7	-0.669	-0.696	-0.540
2014_M	2	-0.760	-0.840	-0.880
2005_CKS	9	-0.647	-0.177	-0.248

CV1: Newly developed lines (50 reps)

	CV1		
	E+L	E+GL+G	E+L+G+GE
AVGDIA	-0.241	0.094	0.124
AVGWT	-0.229	0.159	0.181
BAKEABS	-0.228	-0.057	0.013
BAKETIM	-0.117	0.213	0.224
FLRPRO	-0.242	0.084	0.109
GRNPRO	-0.215	0.160	0.194
IBAKETIM	-0.147	0.220	0.232
IMIXTIM	-0.205	0.235	0.246
LOFVOL	-0.246	0.080	0.134
MIXABS	-0.229	-0.037	0.010
MIXTIM	-0.197	0.234	0.236
TESTWT	-0.232	0.107	0.156

CV0: New environments

LOFVOL

Environment	Sample Size	CV0		
		L+E	L+E+G	L+E+G+GE
2009_CKS	102	-0.060	-0.067	-0.089
2010_CKS	43	0.115	0.050	0.081
2011_CKS	35	0.142	0.279	0.216
2012_BE	65	0.440	0.509	0.496
2012_GY	6	0.861	0.861	0.861
2012_K	6	0.289	0.289	0.289
2012_M	6	0.074	0.074	0.074
2012_MP	66	0.402	0.385	0.411
2008_CKS	68	-0.033	0.010	-0.037
2006_CKS	34	0.450	0.337	0.353
2013_GY	46	0.511	0.497	0.503
2013_K	5	0.495	0.616	0.338
2013_MP	40	0.189	0.172	0.211
2007_CKS	8	-0.706	-0.677	-0.683
2014_CKS	133	0.145	0.020	0.002
2014_SU	7	-0.230	-0.521	-0.498
2014_M	2	-1.000	-1.000	-1.000
2005_CKS	9	0.717	0.757	0.694

CV0: New environments

	CV0		
	E+L	E+GL+G	E+L+G+GE
AVGDIA	0.373	0.412	0.413
AVGWT	0.293	0.379	0.366
BAKEABS	0.110	0.101	0.113
BAKETIM	0.443	0.454	0.451
FLRPRO	0.181	0.252	0.246
GRNPRO	0.221	0.290	0.282
IBAKETIM	0.443	0.469	0.466
IMIXTIM	0.387	0.436	0.438
LOFVOL	0.190	0.168	0.157
MIXABS	0.104	0.113	0.109
MIXTIM	0.407	0.458	0.461
TESTWT	0.278	0.325	0.319

CV00: Forward prediction

LOFVOL

		L + E			E + L +G			E + L + G + GE		
		2011	2012	2013	2011	2012	2013	2011	2012	2013
2006_CKS	34									
2008_CKS	68									
2009_CKS	102									
2010_CKS	43									
2011_CKS	35									
2012_BE	65	-0.027			-0.209			-0.182		
2012_MP	66	0.043			-0.361			-0.347		
2013_GY	46	-0.275	0.150		0.000	0.066		0.108	0.126	
2013_MP	40	-0.132	0.492		-0.078	0.519		0.080	0.546	
2014_CKS	133	0.172	0.032	-0.147	0.261	0.310	0.346	0.253	0.277	0.323
		0.017	0.108	-0.147	-0.016	0.228	0.346	0.020	0.226	0.323

Conclusions

- **For CV1 and CV2.**
  - ❖ Acceptable results for complex traits using the main effects model (E + L +G)
  - ❖ Slight improvement in predictive ability by including the Genotype – by – Environment interaction

- **For CV0.**
  - ❖ Predicting new environments models E + L + G and E + L + G + GE performed the same.

- **For CV00.**
  - ❖ In forward prediction main effects model performed the better.

Thank you!

UNIVERSITY OF Nebraska Lincoln