Genomes to Fields

GxE interaction as a driver for enhance Whole Genomic Prediction on 2014-15 G2F data

Diego Jarquin

December 8, 2016

www.Genomes2Fields.org

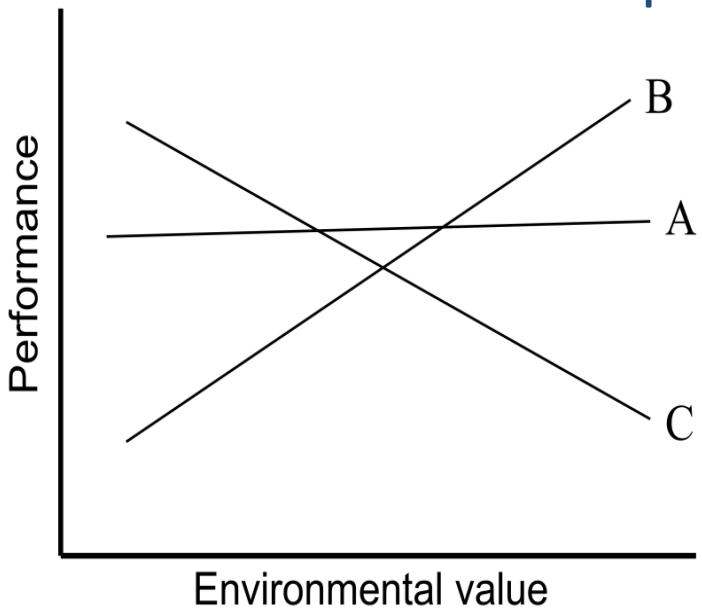
Outline

- ✧Introduction
  - ✧The nature of the problem
  - ✧Genomic selection in a nutshell
- ✧Data description
  - ✧GBS
  - ✧Synthetic genotypes
  - ✧Experimental design
  - ✧Weather records
- ✧Materials and Methods
  - ✧Models
  - ✧Cross-validation Schemes - Predicting traits: from genomes to fields
- ✧Results
  - ✧Predictions
- ✧Discussion, Conclusions, and Future studies

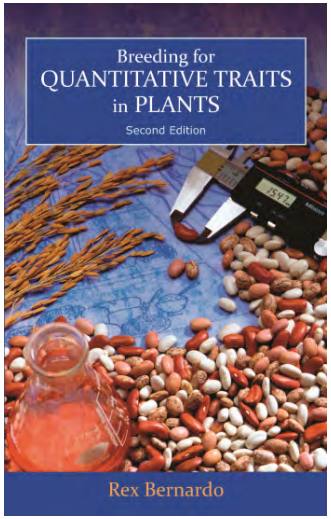
Genomes to Fields

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



Whole Genomic Prediction

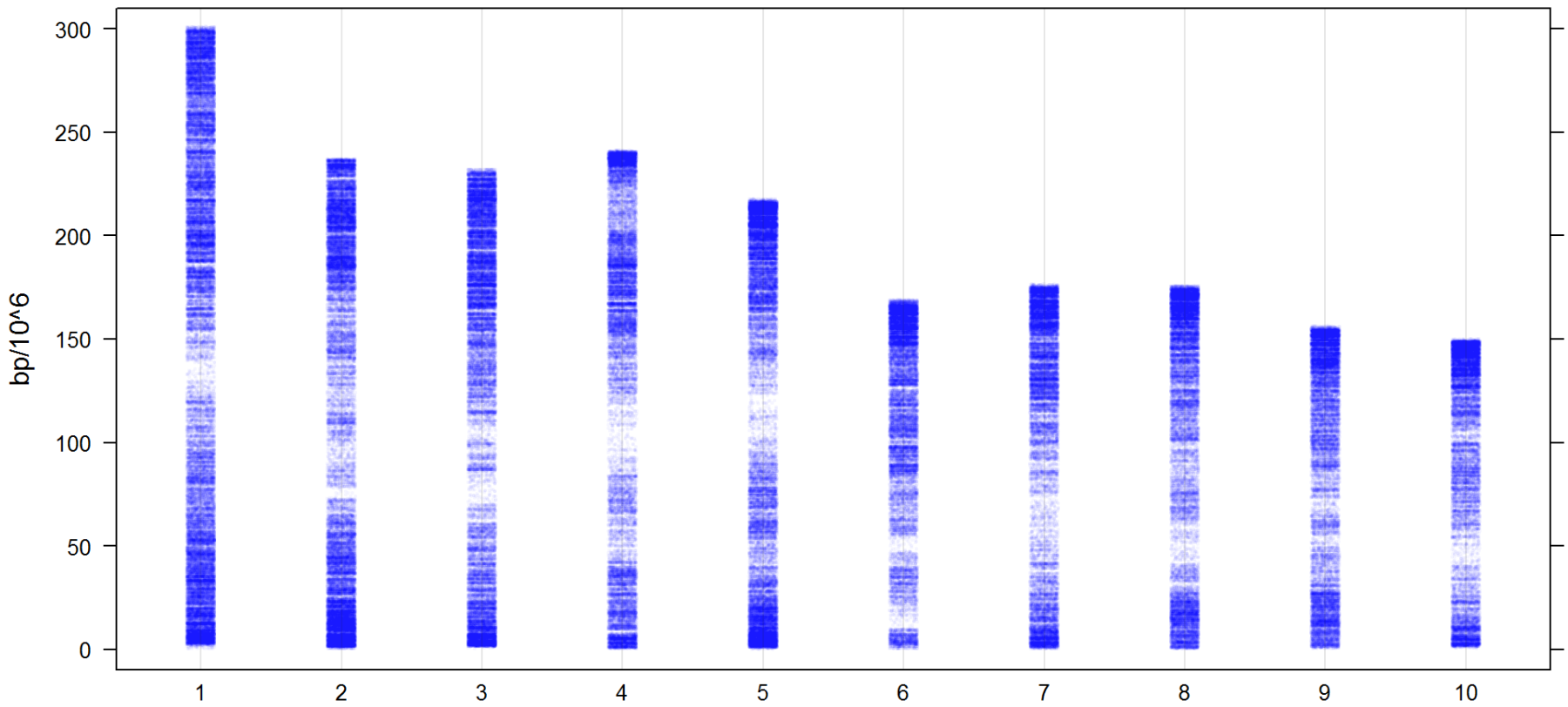
Combines genotypic and phenotypic information to calibrate models and perform prediction on un-phenotyped individuals using molecular markers to stablish genetic relationships with the initial set.

Data description: GBS Data

✧ Close to 1,000,000 SNPs for 543 Inbreds.

757,301 SNPs
20% missing data

245,217 SNPs
20% missing data & MAF < 0.05



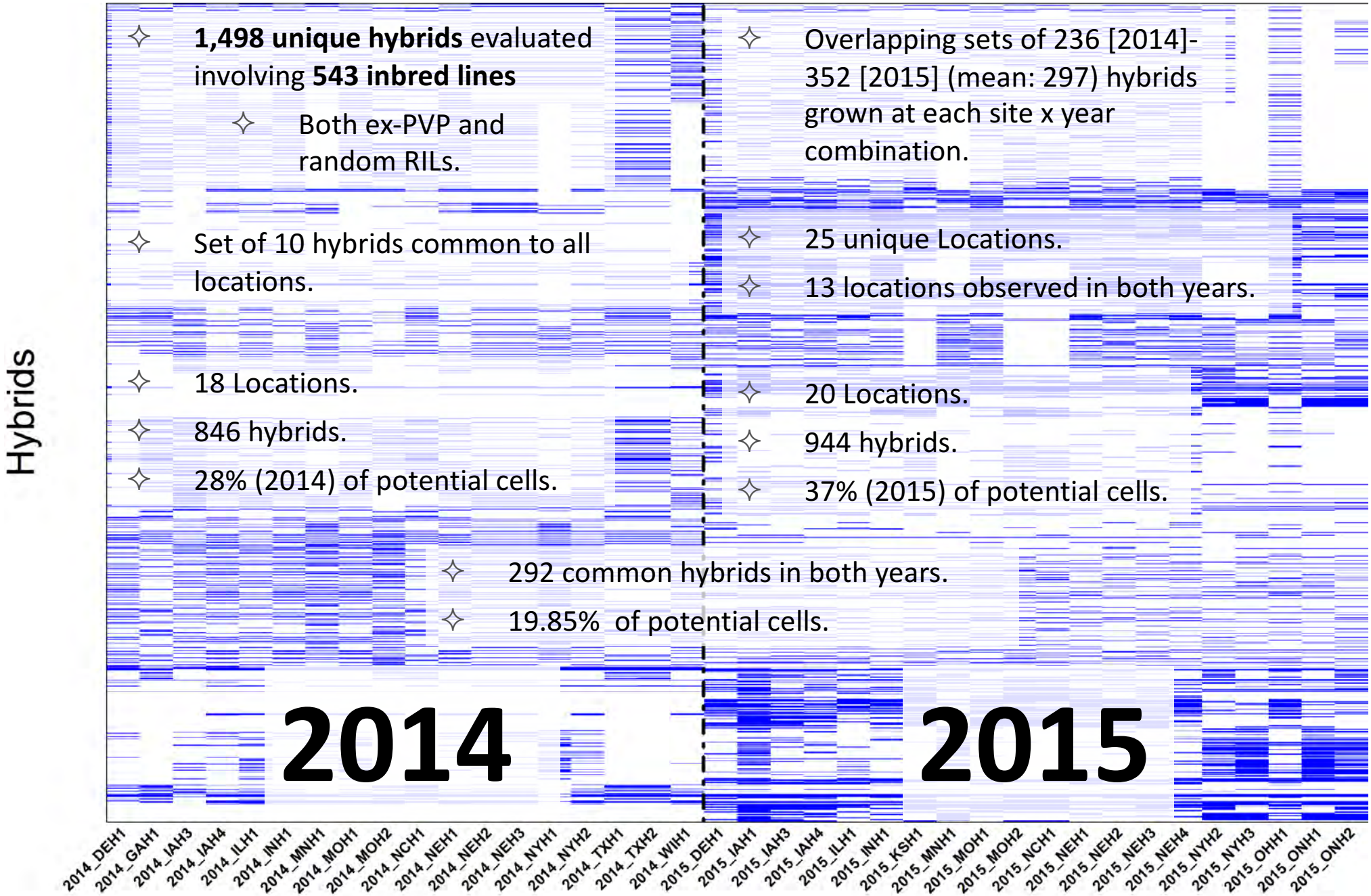
Data description: “Synthetic Genotypes”

- Added together SNP scores of parents to score hybrid genotype

Parent 1		Parent 2		Hybrid
1	+	1	=	1
-1		-1		-1
-1		1		0
1		-1		0
1		-1		0
0.75		1		0.88
-1		-1		-1

Genomes to Fields

Data description: Experimental design (two years)

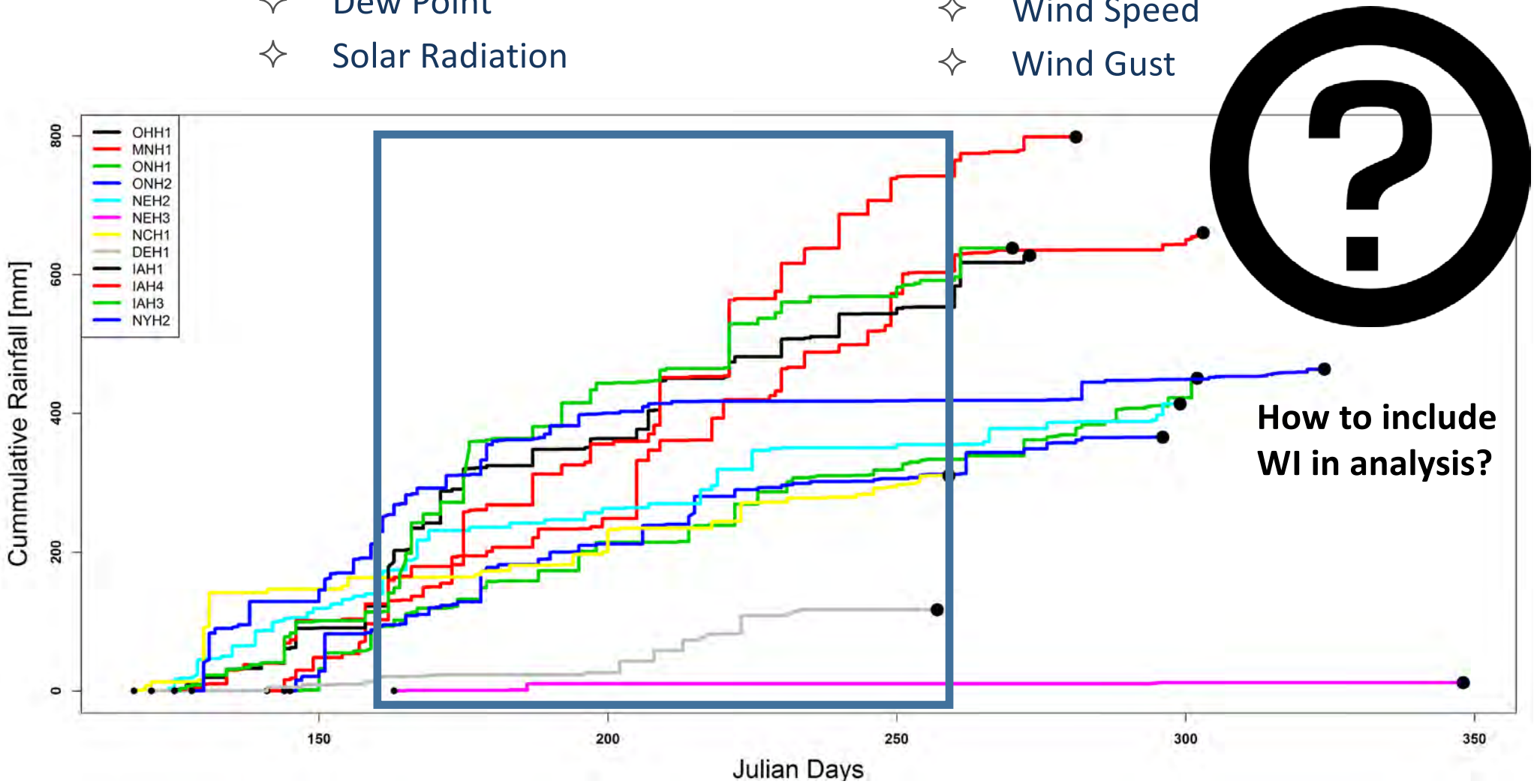


- **1,498 unique hybrids** evaluated involving **543 inbred lines**
  - Both ex-PVP and random RILs.
- Set of 10 hybrids common to all locations.
- 18 Locations.
- 846 hybrids.
- 28% (2014) of potential cells.
- Overlapping sets of 236 [2014]-352 [2015] (mean: 297) hybrids grown at each site x year combination.
- 25 unique Locations.
- 13 locations observed in both years.
- 20 Locations.
- 944 hybrids.
- 37% (2015) of potential cells.
- 292 common hybrids in both years.
- 19.85% of potential cells.

Data description: Weather Information (WI)

8 Environmental Covariates (ECs) hourly recorded.

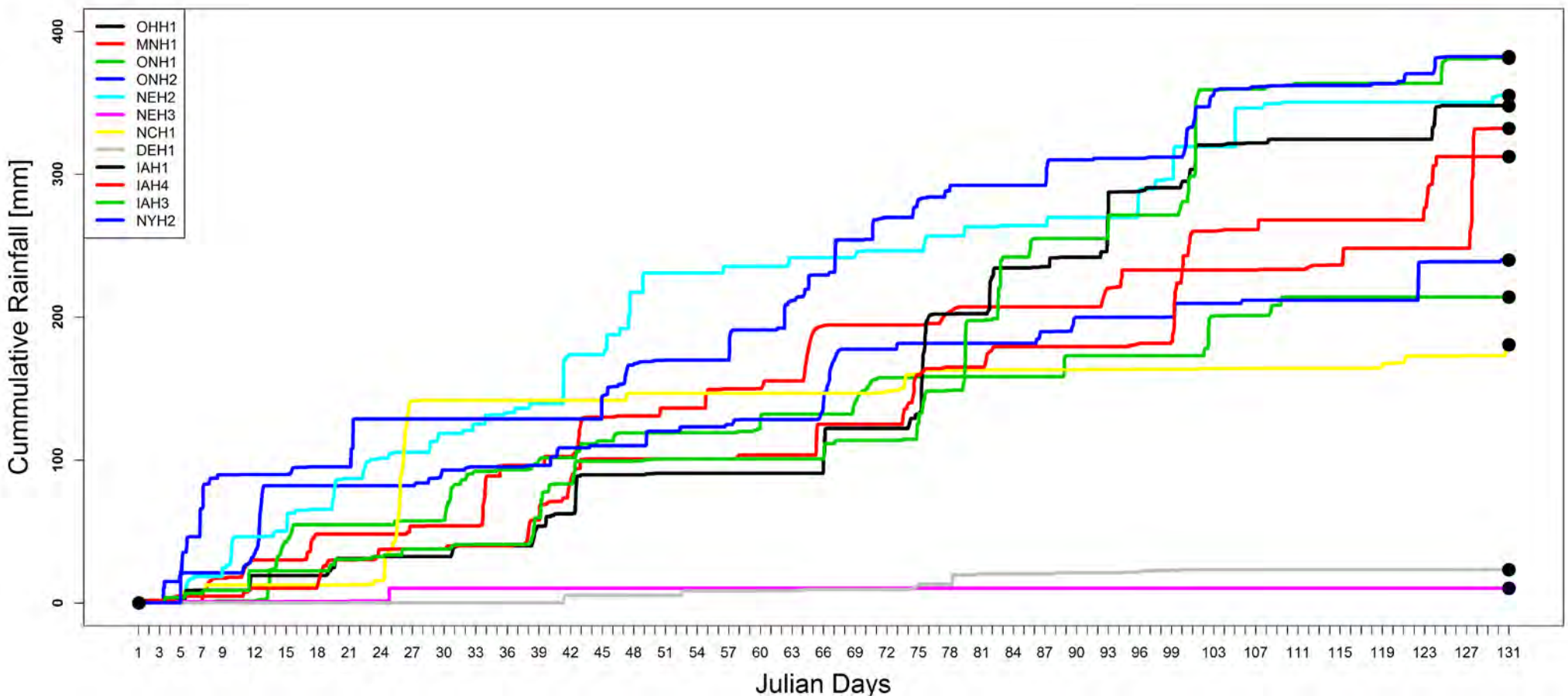
- Temp C*
- Humidity
- Dew Point
- Solar Radiation
- Rain Fall
- Wind Direction
- Wind Speed
- Wind Gust



Data description: Weather Information (WI)

Setting to zero days the planting date 3 methods to include WI were tested.

- 131 plain or absolute days common to all environments.
- 8 (ECs) x 24 (hours) x 131 (days) = 25,152 co-variates per environment (W1).
- Computing min, max, and mean per day for each EC gives 8 (ECs) x 3 (min, max, mean) x 131 (days) = 3,144 covariates per environment (W2).

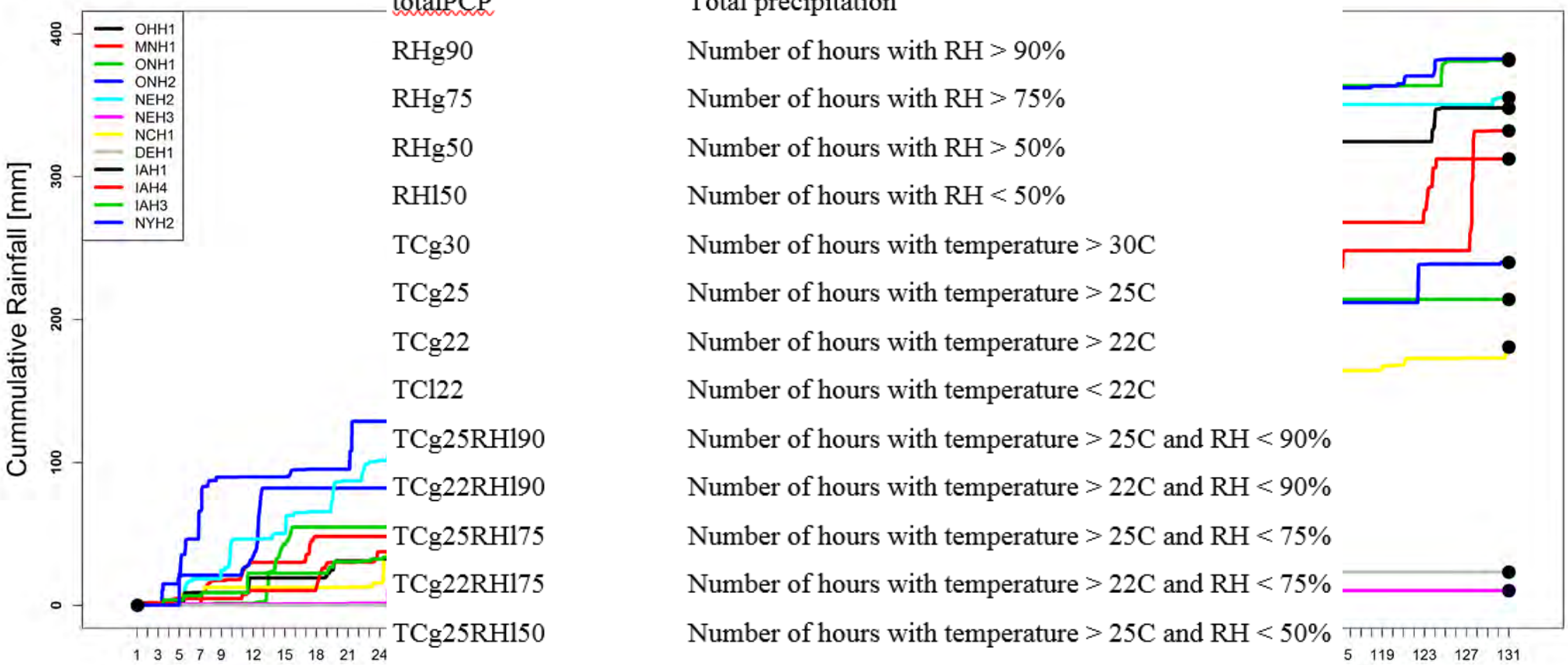


Data description: Weather Information (WI)

Setting to zero days the planting date 3 methods to include WI were tested.

- Four time intervals (1-30, 31-60, 61-90, and 91-131) and 21 ECs per period for a total of 84 ECs (W3) (Perez-Rodriguez et al., 2015):

avgTemp	Average temperature (C)
maxTemp	Maximum temperature (C)
minTemp	Minimum temperature (C)
avgRH	Average relative humidity (%)
totalPCP	Total precipitation
RHg90	Number of hours with RH > 90%
RHg75	Number of hours with RH > 75%
RHg50	Number of hours with RH > 50%
RHl50	Number of hours with RH < 50%
TCg30	Number of hours with temperature > 30C
TCg25	Number of hours with temperature > 25C
TCg22	Number of hours with temperature > 22C
TCl22	Number of hours with temperature < 22C
TCg25RHl90	Number of hours with temperature > 25C and RH < 90%
TCg22RHl90	Number of hours with temperature > 22C and RH < 90%
TCg25RHl75	Number of hours with temperature > 25C and RH < 75%
TCg22RHl75	Number of hours with temperature > 22C and RH < 75%
TCg25RHl50	Number of hours with temperature > 25C and RH < 50%
TCg22RHl50	Number of hours with temperature > 22C and RH < 50%



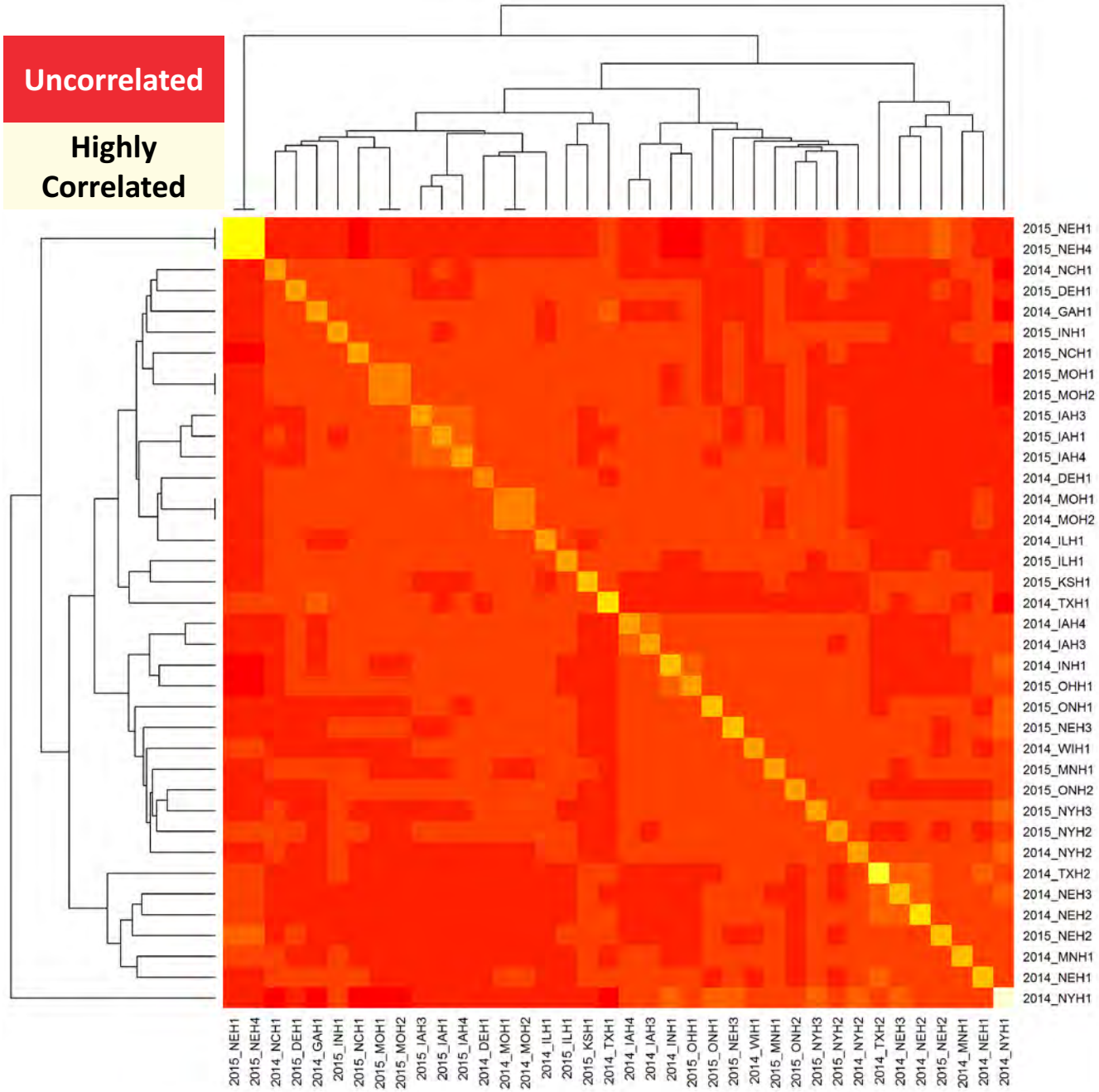
Weather similarities between pairs of Environments using W1 ECs.

Matrix of similarities using:

Hourly records as covariates.
8 x 24 x 131 = 25,152 ECs.

Low values were observed in the off-diagonal entries.

Only a few environments showed moderate correlations.



Weather similarities between pairs of Environments using W2 ECs.

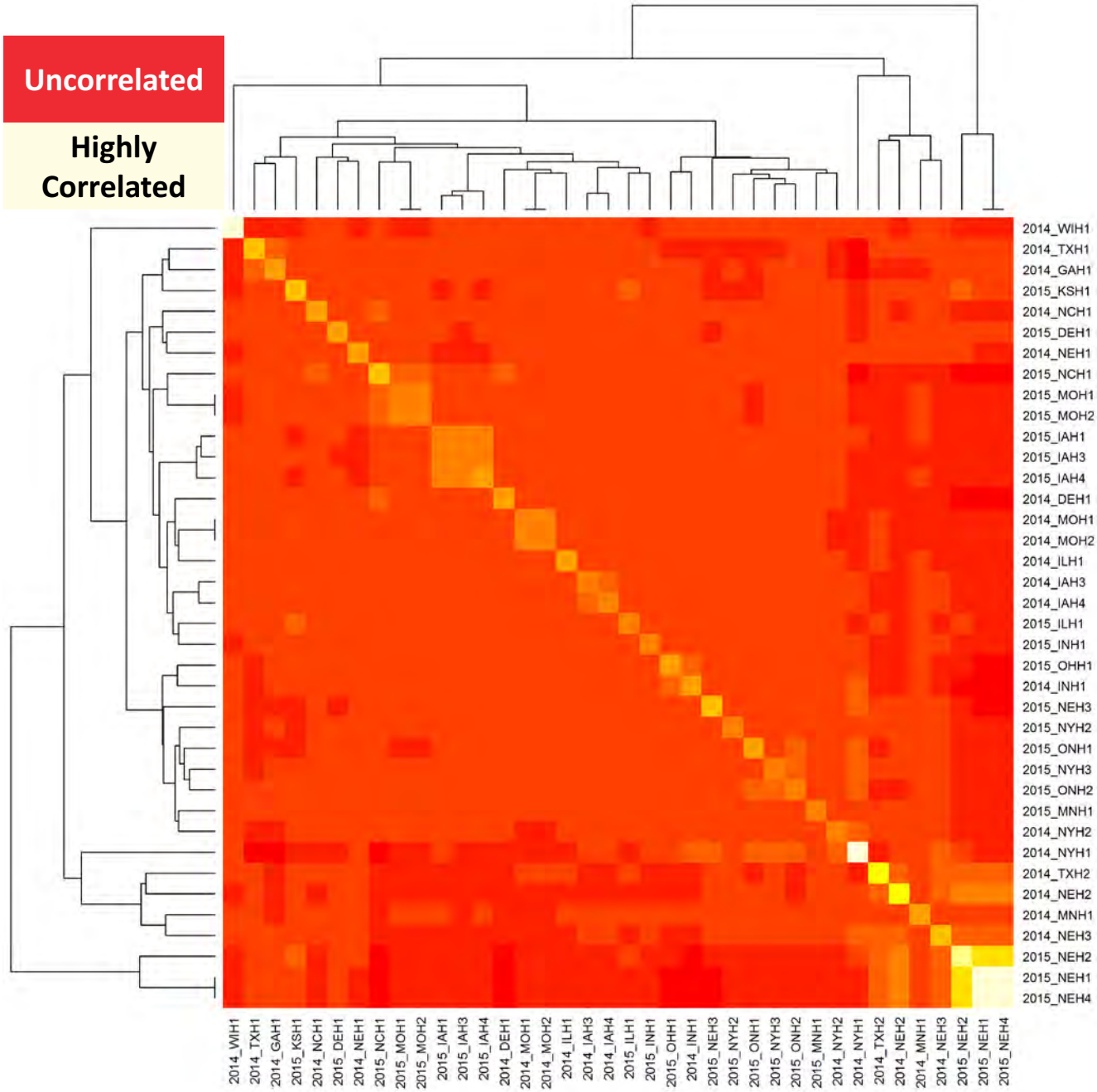
Matrix of similarities using:

Min, max, & means per day.

8 x 3 x 131 = 3,144 ECs.

A slight improvement connecting environments.

But still the improvements look very poor.



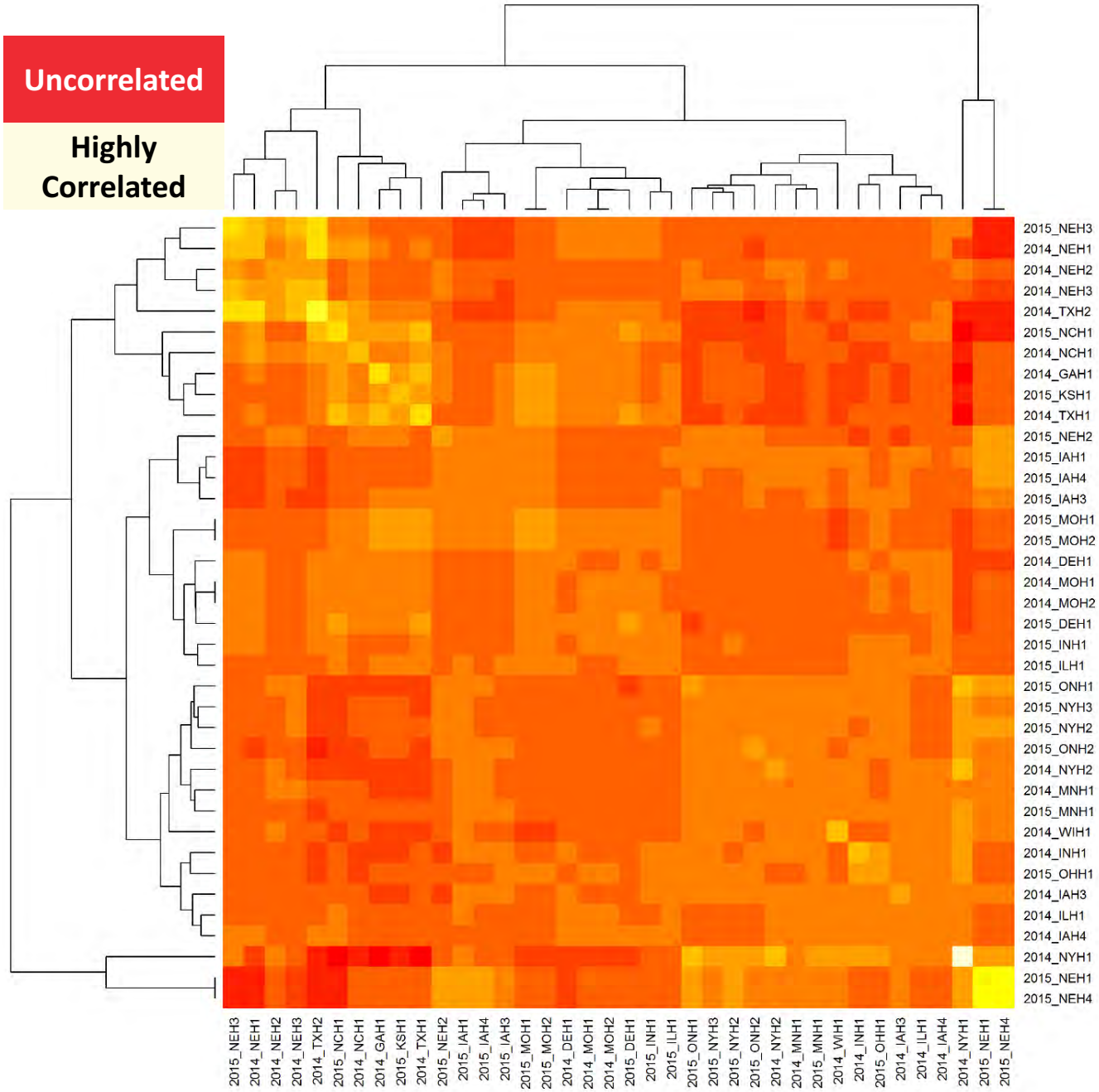
Weather similarities between pairs of Environments using W3 ECs.

Matrix of similarities using:

Means of 21 covariates measured in four different time periods [1-30], [31-60], [61-90], & [91-131] days.

21 x 4 = 84 ECs.

Better connectivity between environments via ECs.



Models based on co-variance structures (Jarquin et al, 2014)

M1: E + L	**Model without marker data.**
M2: E + L + G	**Baseline model (maker data), similar results are expected using other Genomic Prediction models (Bayes Alphabet, Penalized Methods).**
M3: E + L + G + GW	**Model including interactions between markers and environmental covariates.**
M4: E + L + G + GW + GE	**Model accounting for imperfect information via GE (interactions between markers and environments).**

- **All terms were treated as random effects**
  - **E: environment effect.**
  - **L : line effects.**
  - **G: main effect of the markers.**
  - **GW: interactions between makers and environmental co-variates (W1, W2, and W3).**
  - **GE: interactions between markers and environments.**
  - **A total of 8 models were tested (M3 and M4 fitted for each ECs data set [W1, W2, & W3]).**

Variance Components (Evidence of GxE)

Environment component captures around 60% of the phenotypic variability

	Variance components							
Models	E	L	G	GE	GW1	GW2	GW3	R
E+L	1231.6	224.6						595.4
E+L+G	1209.4	88.6	234.6					594.0
E+L+G+GW1	1286.6	79.8	202.5		188.7			462.1
E+L+G+GW2	1283.0	78.8	205.8			194.3		464.5
E+L+G+GW3	1269.5	76.5	221.6				179.6	502.5
E+L+G+GW1+GE	1272.0	80.7	196.7	95.9	116.3			445.9
E+L+G+GW2+GE	1270.1	80.2	199.7	104.5		109.5		446.0
E+L+G+GW3+GE	1276.9	79.3	199.0	137.3			81.7	447.3

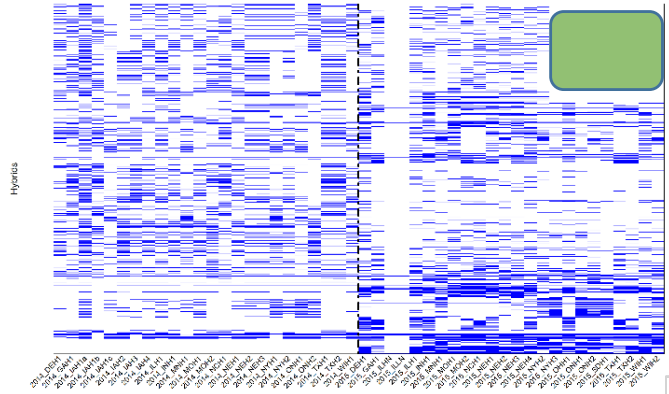
Interactions explains a sizable proportion of the variability.

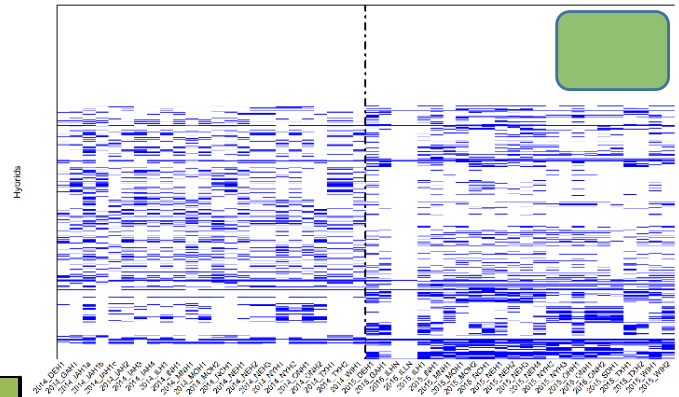
Almost the same amount is explained by markers.

Two types of interactions capture more variability than just one.

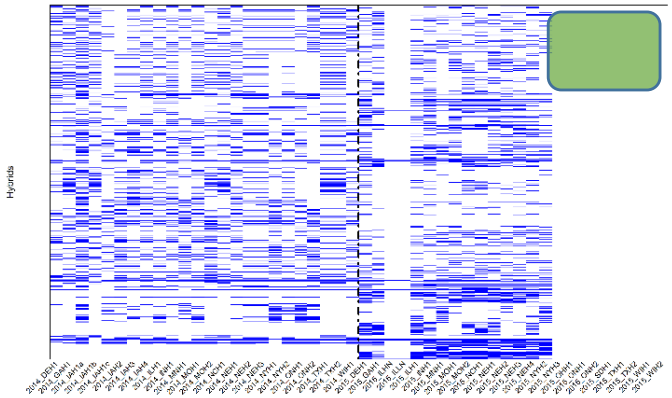
	Within environment variability						
Models	L	G	GE	GW1	GW2	GW3	R
E+L	27.4						72.6
E+L+G	9.7	25.6					64.8
E+L+G+GW1	8.6	21.7		20.2			49.5
E+L+G+GW2	8.4	21.8			20.6		49.2
E+L+G+GW3	7.8	22.6				18.3	51.3
E+L+G+GW1+GE	8.6	21.0	10.2	12.4			47.7
E+L+G+GW2+GE	8.5	21.2	11.1		11.7		47.5
E+L+G+GW3+GE	8.4	21.1	14.5			8.7	47.3

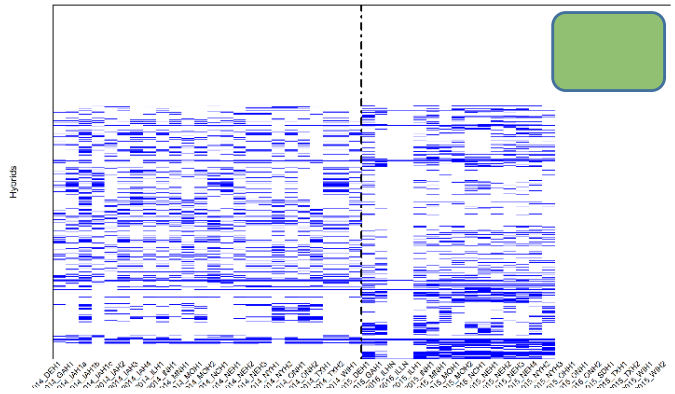
Predicting traits: from genomes to fields





		Tested Genotypes	
		YES	NO
Tested Environments	YES	CV2	CV1
	NO	CV0	CV00





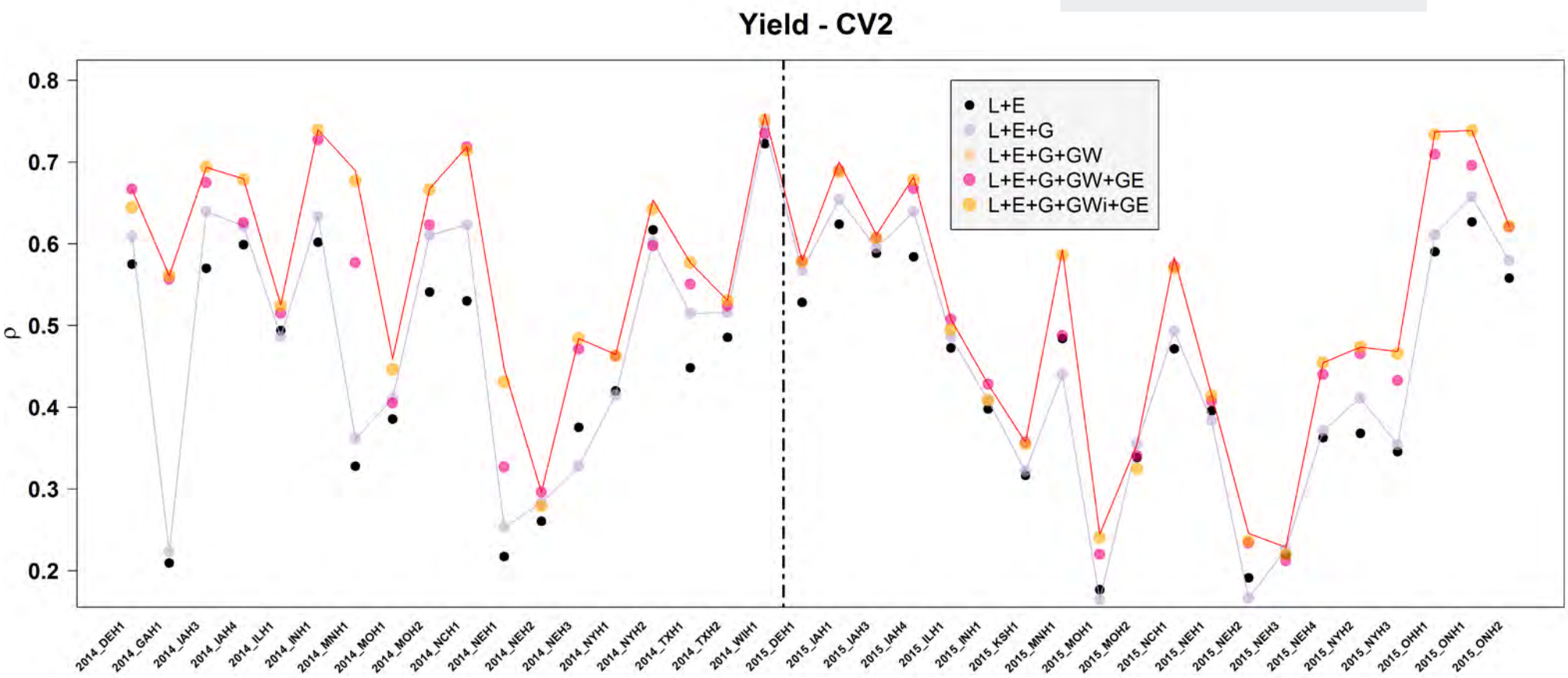
Genomes to Fields

Results: Predictions – 80% training and 20% testing

		Tested Genotypes	
		YES	NO
Tested Environments	YES	CV2	CV1
	NO	CV0	CV00

L+E	0.44
L+E+G	0.46
L+E+G+GW	0.52
L+E+G+GW+GE	0.54
L+E+G+GWi+GE	0.55

17% Improvement

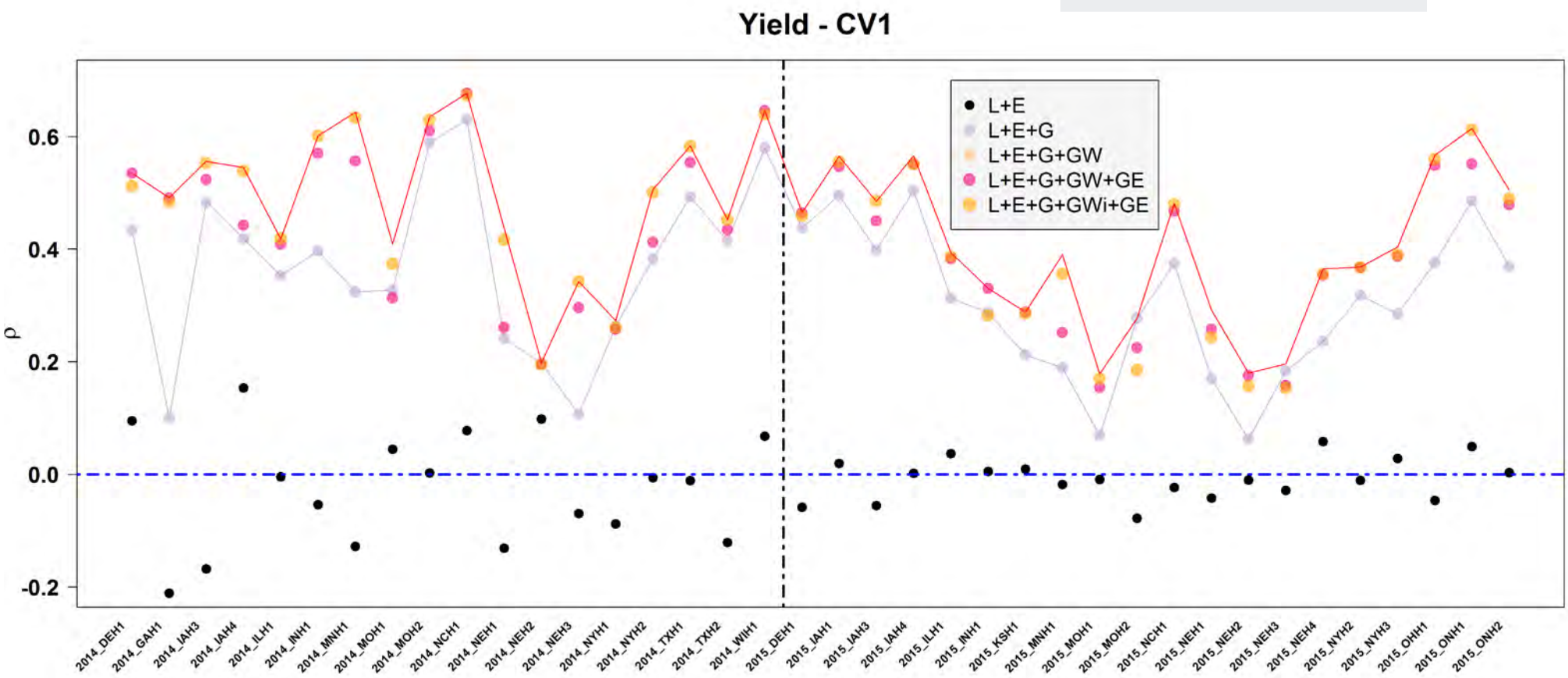


Results: Predictions – 80% training and 20% testing

		Tested Genotypes	
		YES	NO
Tested Environments	YES	CV2	CV1
	NO	CV0	CV00

L+E	0.00
L+E+G	0.34
L+E+G+GW	0.41
L+W+G+GW+GE	0.43
L+W+G+GWi+GE	0.44

32% Improvement

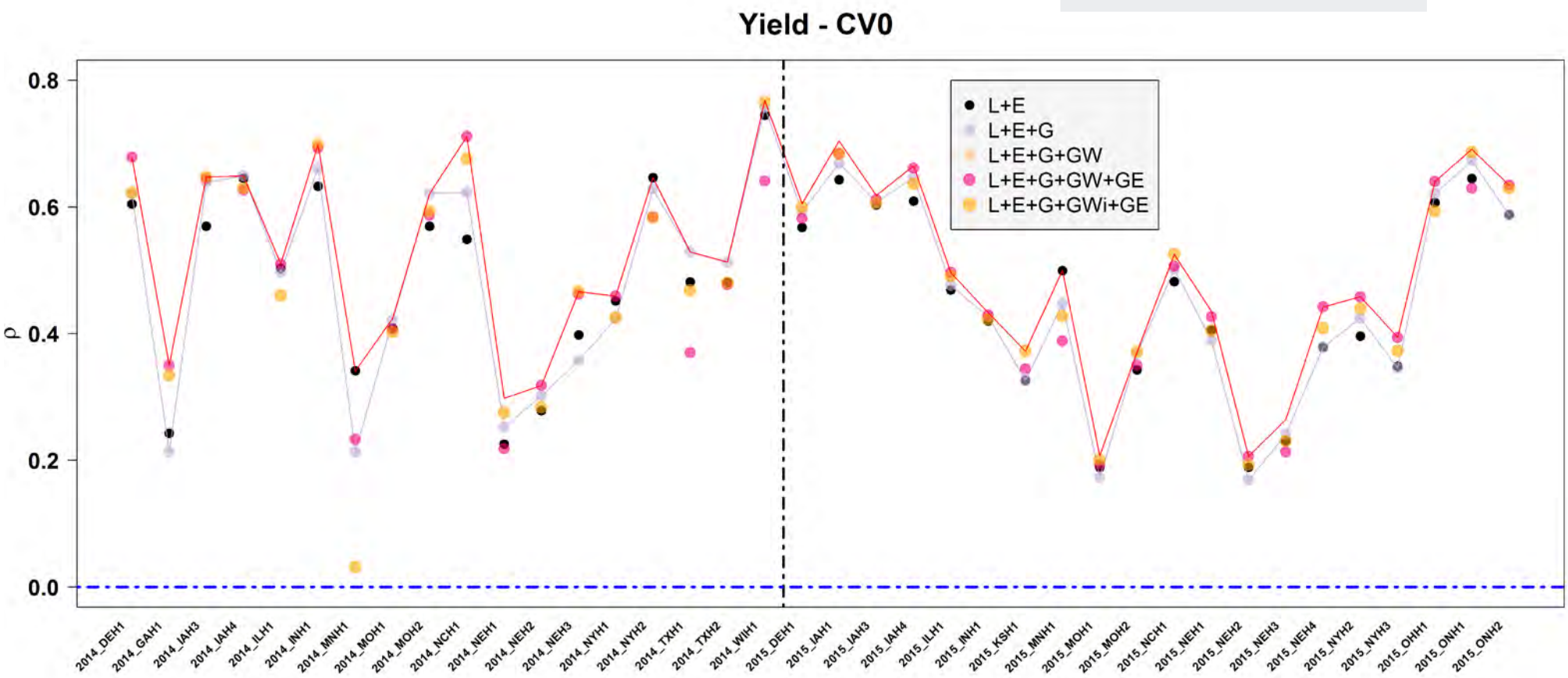


Results: Predictions – one location at a time including replicates

		Tested Genotypes	
		YES	NO
Tested Environments	YES	CV2	CV1
	NO	CV0	CV00

L+E	0.46
L+E+G	0.47
L+E+G+GW	0.48
L+E+G+GW+GE	0.47
L+E+G+GWi+GE	0.51

7% Improvement

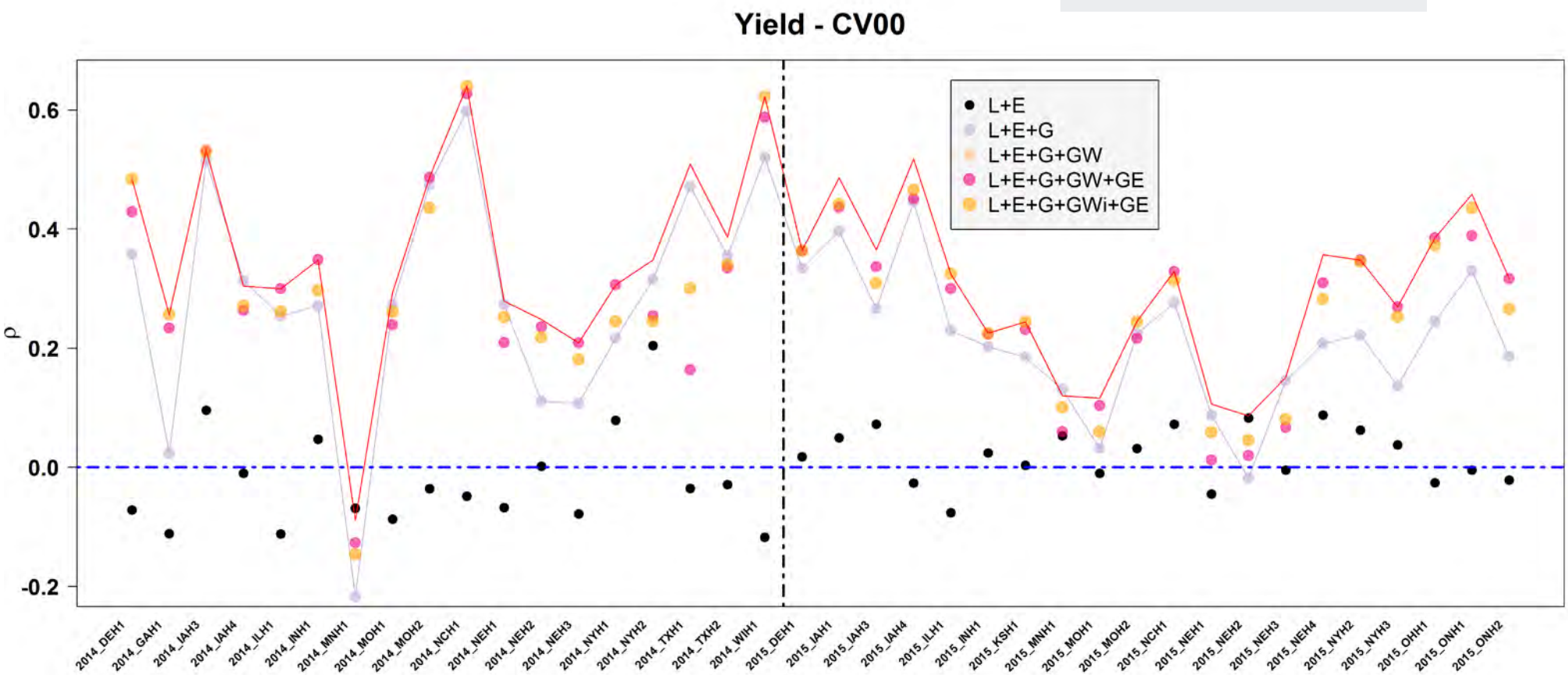


Results: Predictions – one location at a time without replicates

		Tested Genotypes	
		YES	NO
Tested Environments	YES	CV2	CV1
	NO	CV0	CV00

L+E	0.00
L+E+G	0.25
L+E+G+GW	0.28
L+E+G+GW+GE	0.29
L+E+G+GWi+GE	0.32

30% Improvement



Discussion and conclusions

- **Different prediction problems gave different results.**

- **Interaction components account for close to 25% of within environment variability.**

- **Interaction models work well in all schemes**
  - Sizable improvements in predictive ability [7-32%] with respect to the baseline model.

- **With interaction models good results predicting 20% of missing are expected.**
  - Incomplete field trials scenario [CV2] showed an average Predictive Ability (aPA) of 0.55; 17% improvement with respect to baseline model.
  - Newly developed lines [CV1] gives an aPA of 0.44; 32% improvement.

Genomes to Fields

Discussion and conclusions

✧Predicting new environments

- ✧The different ECs co-variance matrices improved PA in most of the cases; however, there was not a unique co-variance structure outperforming the others.

- ✧Including replicates observed in other environments [CV0] genomic information lose relevance (all models perform similarly).

- ✧ However, ECs might improve aPA (0.51) about 7% (we will work to explain why and how to take advantage of this).

Genomes to Fields

Discussion and conclusions

- **Predicting unobserved genotypes in unobserved environments (CV00) genomic information becomes the main source of information.**
  - The aPA was about 0.32 selecting the best results according to the different models (30% improvement).

- **Future studies.**
  - Include ECs in a more informed way to connect observed and unobserved sites with important maize stages.
  - Incorporate information from other sources (aerial images).

Genomes to Fields

GxE Consortium: Data Usage Disclaimer

This presentation includes data analysis and interpretation conducted by the presenter and does not necessarily reflect the observations and conclusions of the GxE Consortium.

Genomes to Fields

G X E Cooperators

Principal Investigators who grew GxE trials in 2014-2016

- Martin Bohn (UIUC)
- Ed Buckler (ARS)
- Ignacio Ciampitti (KSU)
- Jode Edwards (ARS)
- Sherry Flint-Garcia (ARS)
- Mike Gore (Cornell)
- Christopher Graham (SDSU)
- Candy Hirsch (UMN)
- Jim Holland (ARS)
- Elizabeth Hood (A-State)
- David Hooker (Guelph)
- Fiona Goggin (A-State)
- Shawn Kaeppler (UW)
- Joe Knoll (ARS)
- Judith Kolkman (Cornell)
- Greg Kruger (UNL)
- Nick Lauter (ARS)
- Liz Lee (Guelph)
- Natalia de Leon (UW)
- Argelia Lorence (A-State)
- Aaron Lorenz (UMN)
- Jonathan Lynch (PSU)
- Steve Moose (UIUC)
- Seth Murray (TAMU)
- Rebecca Nelson (Cornell)
- Torbert Rocheford (Purdue)
- Oscar Rodriguez (UNL)
- Cinta Romay (Cornell)
- James Schnable (UNL)
- Brian Scully (ARS)
- Rajandeep Sekhon (Clemson)
- Margaret Smith (Cornell)
- Nathan Springer (UMN)
- Kurt Thelen (MSU)
- Peter Thomison (OSU)
- Mitch Tuinstra (Purdue)
- Jason Wallace (UGA)
- Randy Wisser (UDel)
- Wenwei Xu (TAMU)

The Ohio State University
College of Food, Agricultural, and Environmental Sciences
The University of Georgia
United States Department of Agriculture
Agricultural Research Service
the in-house research arm of the USDA
Purdue University
Cornell University
Iowa State University
Clemson University
Illinois
University of Illinois at Urbana-Champaign
Wisconsin
University of Wisconsin–Madison
Michigan State University
University of Minnesota
State Arkansas State University
South Dakota State University
SDSU
University of Nebraska Lincoln
K-State Research and Extension
University of Delaware
Genomes to Fields
University of Guelph
PennState
Texas A&M University

Genomes To Fields Collaborators

† G2F Executive Committee members *GxE Coordinating Groups ˜G2F co-lead

- Martin Bohn (UIUC)*
- Ed Buckler (ARS)†
- Darwin Campbell (ISU)*
- Ignacio Ciampitti (KSU)
- James Clohessy (Cornell)
- Michael Coen (UW)
- Liang Dong (ISU)
- Jode Edwards (ARS)*
- David Ertl (IA Corn)*†
- Sherry Flint-Garcia (ARS)*
- Joseph Gage (UW)
- Jack Gardiner (ISU)*
- Fiona Goggin (A-State)
- Byron Good (Guelph)
- Mike Gore (Cornell)
- Christopher Graham (SDSU)
- Patricio Grassini (UNL)
- Jerry Hatfield (ARS)
- Candy Hirsch (UMN)
- Jim Holland (ARS)
- Elizabeth Hood (A-State)
- David Hooker (Guelph)
- Diego Jarquin (UNL)*
- Shawn Kaeppler (UW)†
- Joe Knoll (ARS)
- Judith Kolkman (Cornell)
- Greg Kruger (UNL)
- Nick Lauter (ARS)
- Carolyn Lawrence-Dill (ISU)*
- Liz Lee (Guelph)*
- Zhizhai Liu (TAMU)
- Natalia de Leon (UW)*†~
- Argelia Lorence (A-State)
- Aaron Lorenz (UMN)*
- Jonathan Lynch (PSU)†
- Nathan Miller (UW)
- Steve Moose (UIUC)
- Seth Murray (TAMU)*
- Rebecca Nelson (Cornell)
- Jane Petzoldt (UW)*
- Torbert Rocheford (Purdue)
- Oscar Rodriguez (UNL)
- Cinta Romay (Cornell)*
- James Schnable (UNL)
- Pat Schnable (ISU)†~
- Brian Scully (ARS)
- Rajandeep Sekhon (Clemson)
- Margaret Smith (Cornell)
- Edgar Spalding (UW)
- Nathan Springer (UMN)†
- Srikant Srinivasan (ISU)*
- Yiwei Sun (ISU)*
- Kurt Thelen (MSU)
- Peter Thomison (OSU)
- Kelly Thorp (ARS)
- Mitch Tuinstra (Purdue)
- Jason Wallace (UGA)
- Rod Williamson (IA Corn)
- Randy Wisser (UDel)*
- Wenwei Xu (TAMU)
- Jianming Yu (ISU)

The Ohio State University
College of Food, Agricultural, and Environmental Sciences

The University of Georgia

United States Department of Agriculture
Agricultural Research Service
the in-house research arm of the USDA

Purdue University

Cornell University

Iowa State University

Clemson University

Illinois
University of Illinois at Urbana-Champaign

Wisconsin
University of Wisconsin–Madison

Michigan State University

University of Minnesota

State
Arkansas State University

South Dakota State University
SDSU

University of Nebraska Lincoln

K-State Research and Extension

University of Delaware

Genomes to Fields

University of Guelph

PennState

Texas A&M University

Genomes To Fields Sponsors

IOWA CORN™ PROMOTION BOARD

NATIONAL CORN GROWERS ASSOCIATION®

Nebraska CornBoard

Arkansas Corn & Grain Sorghum Board

MinnesotaCorn RESEARCH & PROMOTION COUNCIL

Illinois Corn Marketing Board

WISCONSIN CORN PROMOTION BOARD, INC.

KANSAS CORN COMMISSION

TEXAS CORN PRODUCERS

OHIO CORN MARKETING PROGRAM

Genomes to Fields