Integrated Breeding Platform (IBP)

Genotype by Environment Analysis

BMS 13.0-14.0 Manual

- Introduction
- Select Data from Database
- Generate BV Input Files from BMS
- Load Project & Data
- Run Analysis
- Analysis Report & Graphs
  - Descriptive Statistics
    - Trait Summary Statistics
    - Best Variance-Covariance Model for Each Trait
  - Genotype By Environment (GxE) Interactions
    - Stability Superiority Measure
    - Static Stability Measures Coefficients
    - Wrick's Ecovalence Stability Coefficients
    - Finlay and Wilkinson Modified Joint Regression Analysis
    - AMMI Model
    - GGE Model
    - Variance-Covariance Model & Correlation Matrix
    - Correlation Heat Map
    - Scatter Plot Matrix
- References

Introduction

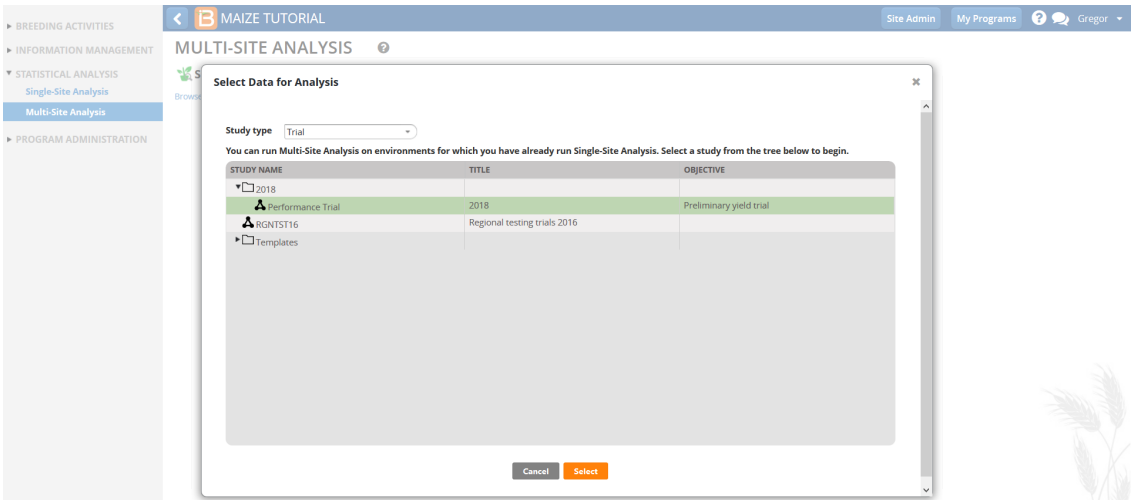
These instructions describes a genotype by environment (GxE) analysis for a four location maize field trial. This tutorial builds upon the adjusted means (BLUEs) and summary statistics calculated for the individual locations (Single Site Analysis).

Select Data from Database

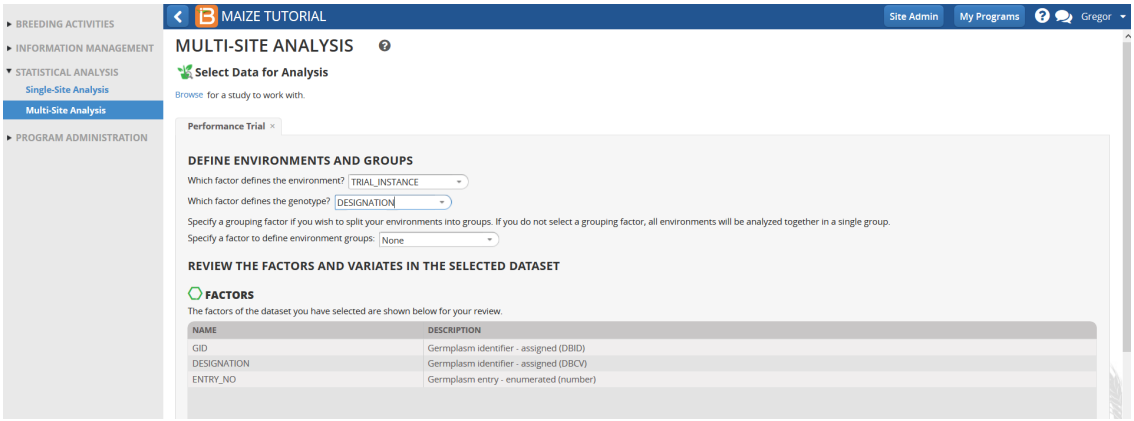
- Open Multi-Site Analysis from the Statistical Analysis menu of the Workbench. Select Browse.



- Select a study with previously uploaded BLUEs and summary statistics from the the single site analysis.

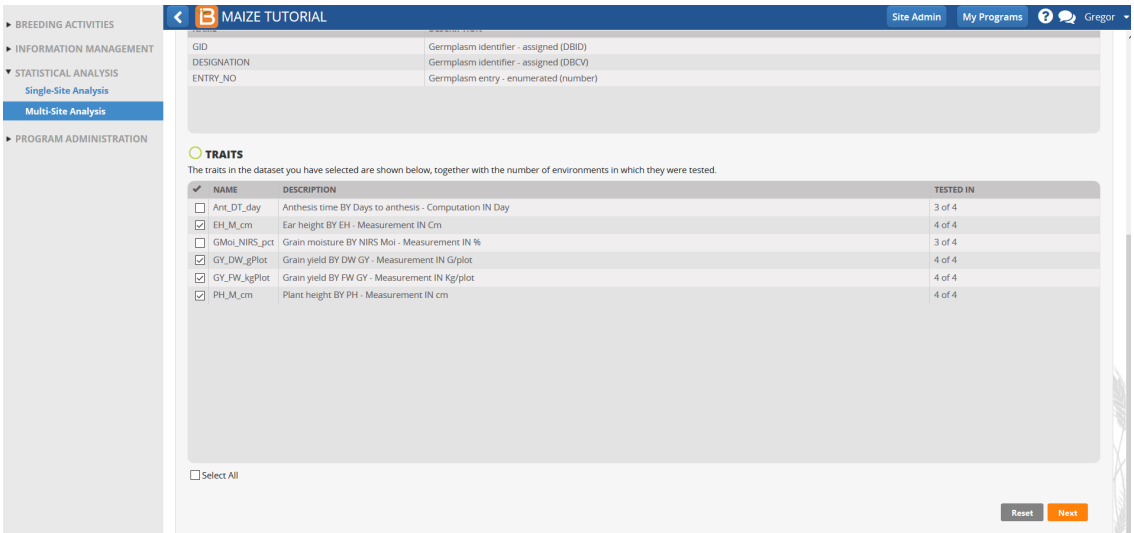


- **Define environments and groups**.
  - Environments: TRIAL_INSTANCE
  - Genotype: DESIGNATION
  - Environment Grouping Factor: None



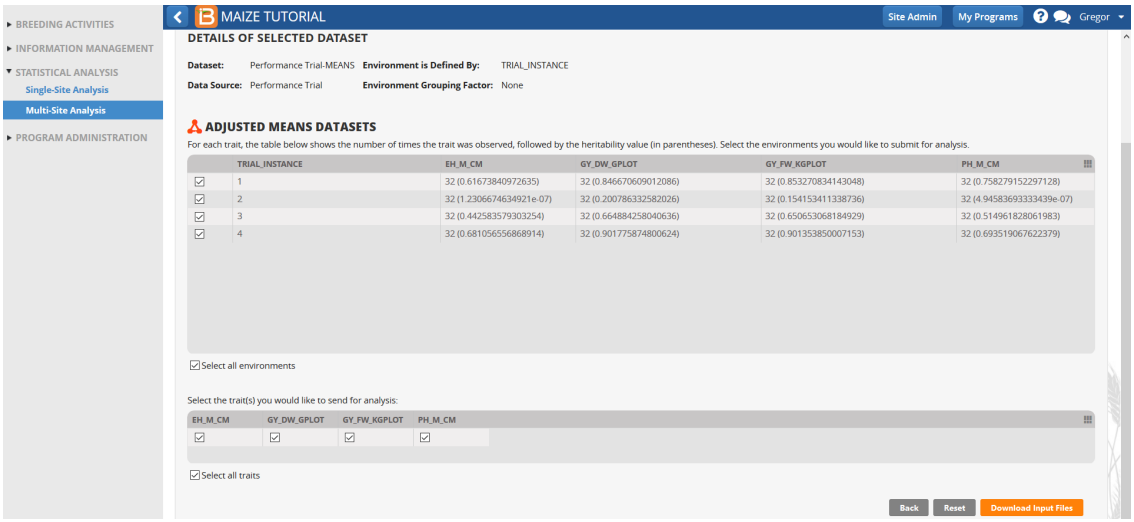
Traits with means available from all trial locations are selected by default. Traits that are not observed or could not be fitted with a mixed model in more than one environment in the single site analysis are not selected for Multi-Site analysis.

- Review the factors and variables in the dataset and select Next.



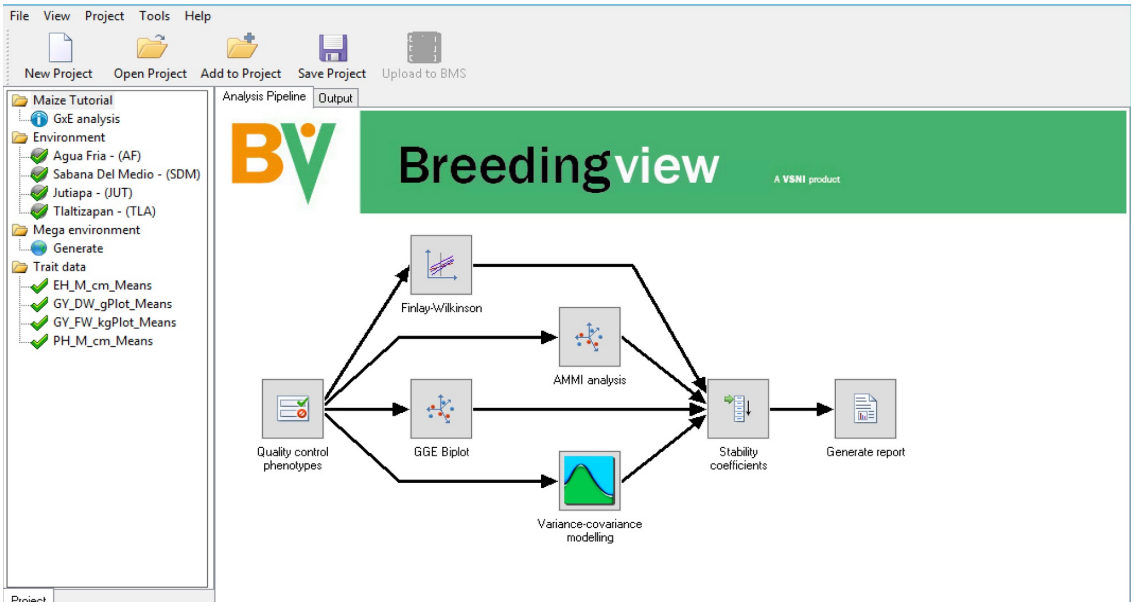
Generate BV Input Files from BMS

- Review the four environments and traits to be included in the multi-site analysis. Select Download Input Files.



Load Project & Data

- The BV Input Files are located within a compressed folder automatically titled Performance Trial. The Breeding View .xml file is located within the Performance Trial folder. Open BV application and select Open Project and browse to the .xml Breeding View project file. The .xml file will load the genotypic and environmental summary statistics.



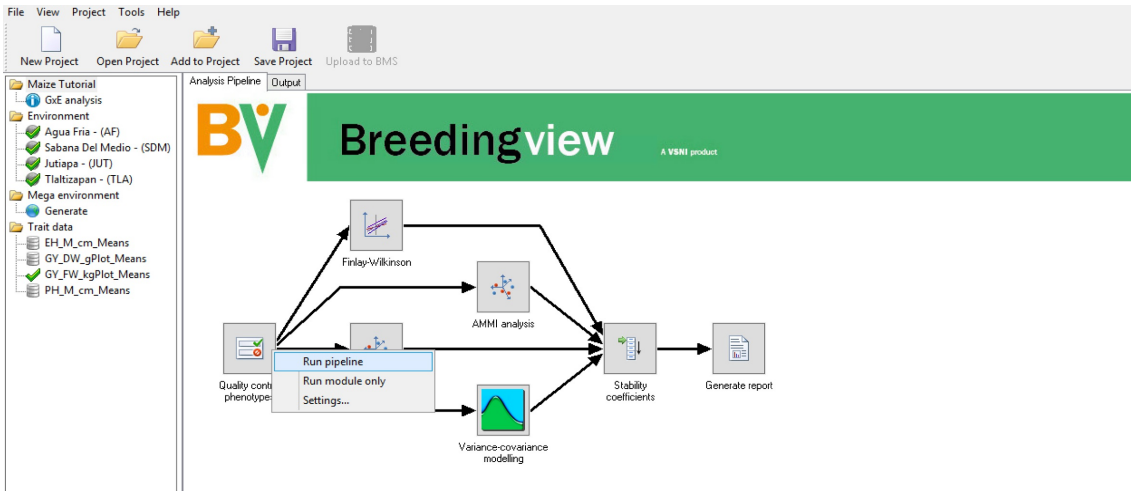
Run Analysis

When a project has been created or opened, a visual representation of the analytical pipeline is displayed in the Analysis Pipeline tab. The analysis pipeline includes a set of connected nodes, which can be used to run and configure pipelines.

Node Descriptions:

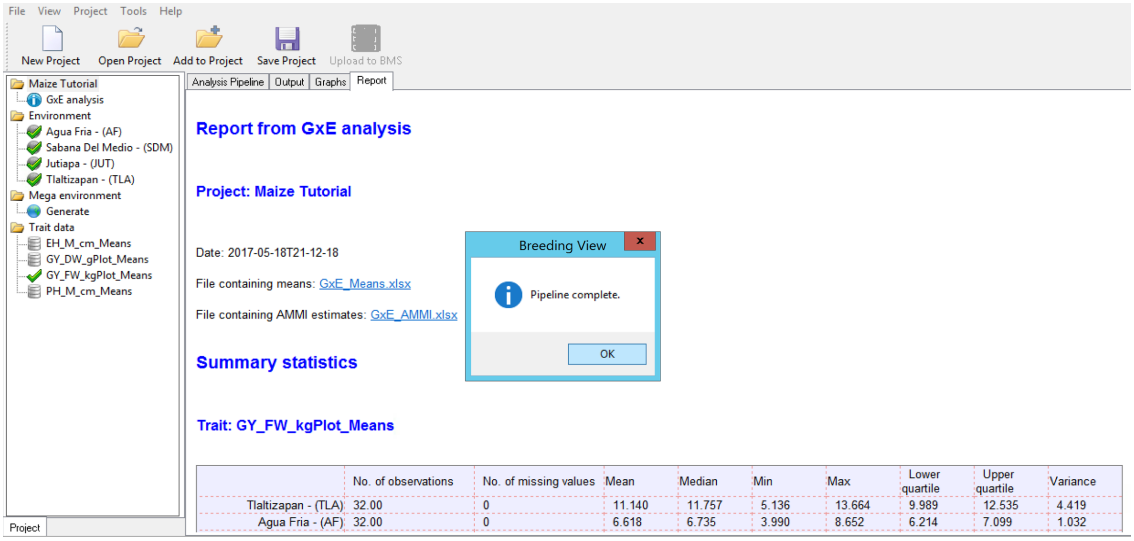
- Quality Control Phenotypes: Summary statistics within and between environments for the trait(s)
- Finlay-Wilkinson: Performs a Finlay-Wilkinson joint regression (Finlay and Wilkinson, 1963)
- AMMI Analysis: Fits an AMMI model and generates summaries and a biplot (Gauch, 1988)
- GGE Biplot: Fits a GGE model and generates a biplot (Yan et al., 2000).
- Variance-Covariance Modeling: Fits different variance-covariance models to the GxE data and selects the best one for the data
- Stability Coefficients: Estimates different stability coefficient parameters to assess genotype performance
- Generate report: Generates an HTML report of the results

- Exclude all traits except grain yield t/ha (GY_FW_kgPlot_Means) from the analysis.
- Run the analysis using the default settings by right clicking the Quality Control Phenotype node and choosing Run Pipeline.



When the analysis is complete a popup notifies the user.

- Select OK.



	No. of observations	No. of missing values	Mean	Median	Min	Max	Lower quartile	Upper quartile	Variance
Tlaltizapan - (TLA)	32.00	0	11.140	11.757	5.136	13.664	9.989	12.535	4.419
Agua Fria - (AF)	32.00	0	6.618	6.735	3.990	8.652	6.214	7.099	1.032

Analysis Report & Graphs

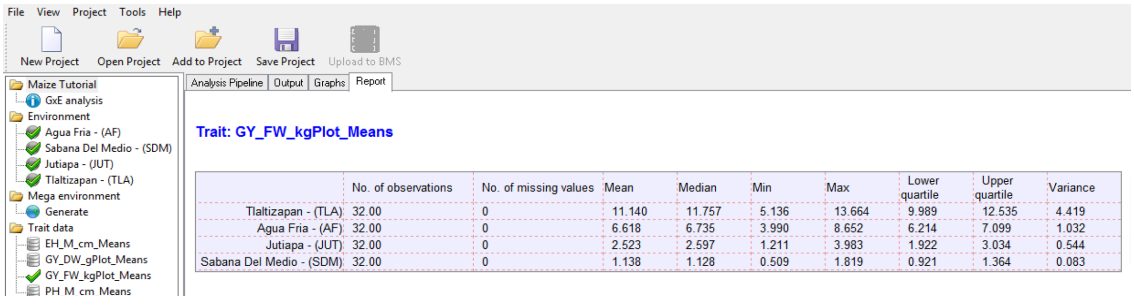
The analysis output can be viewed from Breeding View interface under the results and graphs tabs. Analysis results can also be reviewed as individual files are automatically saved in the location specified by your browser settings, generally the Downloads folder..

Descriptive Statistics

Breeding View provides descriptive statistics that describe the variance and covariance of the entire dataset.

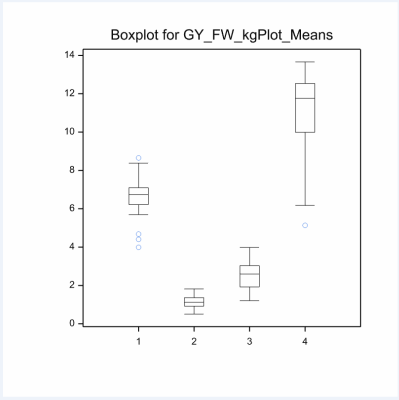
Trait Summary Statistics

The trait summary statistics describe each trait based on the means calculated for each environment in the single site analysis.



	No. of observations	No. of missing values	Mean	Median	Min	Max	Lower quartile	Upper quartile	Variance
Tlaltizapan - (TLA)	32.00	0	11.140	11.757	5.136	13.664	9.989	12.535	4.419
Agua Fria - (AF)	32.00	0	6.618	6.735	3.990	8.652	6.214	7.099	1.032
Jutiapa - (JUT)	32.00	0	2.523	2.597	1.211	3.983	1.922	3.034	0.544
Sabana Del Medio - (SDM)	32.00	0	1.138	1.128	0.509	1.819	0.921	1.364	0.083

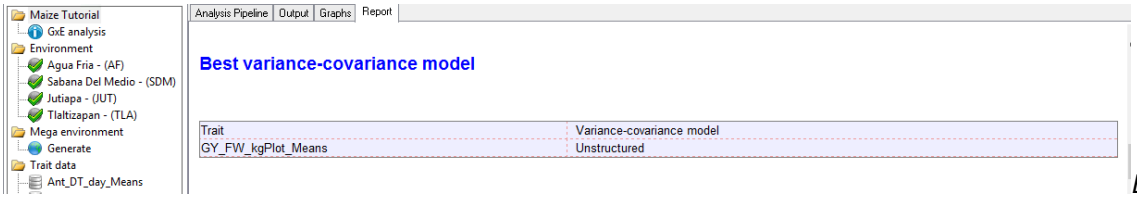
The box plot of means provides a visual representation of the summary statistics.



Boxplot of Grain Yield Means: Tlalzipan(4) has the highest grain yield and the highest variance. Sabana Del Medio (2) has the lowest grain yield and the lowest variance.

Best Variance-Covariance Model for Each Trait

The GxE analysis pipeline formally models the variance-covariance structure in the means data and selects the best model for each trait. The main purpose is to establish a model for later testing of fixed effects, like determining marker effects in a quantitative trait loci by environment (QTLxE) analysis using BLUPs calculated in the single site analysis.



Best Variance-Covariance Model: In this example, grain yield means are best described by an unstructured model, where each variance and covariance is estimated uniquely from the data.?

Genotype By Environment (GxE) Interactions

Stability, or lack of phenotypic plasticity, is calculated for each genotype considering all traits using the following analyses:

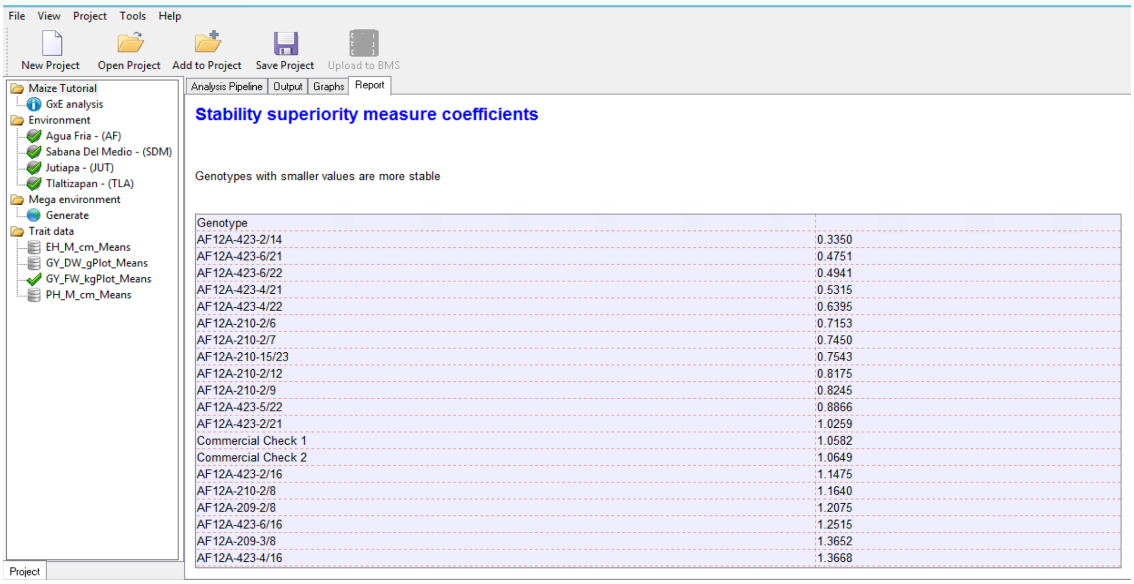
- Cultivar-Superiority Measure
- Static Stability Measures Coefficients
- Wricke's Ecovalence Stability Coefficients

GxE interactions are also examined for each individual trait using the following analyses:

- Finlay and Wilkinson Modified Joint Regression
- AMMI Model
- GGE Model
- Best Variance-Covariance Model
- Correlation Matrix
- Scatter Plot Matrix

Stability Superiority Measure

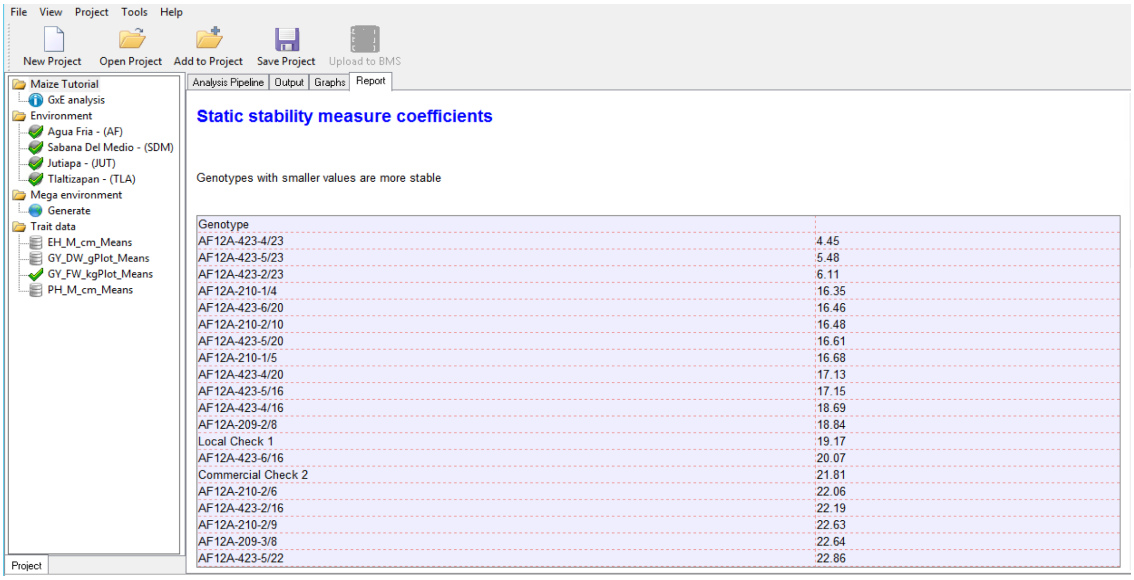
Stability Superiority Measure (Lin & Binns,1988) is the sum of the squares of the difference between genotypic mean in each environment and the mean of the best genotype, divided by twice the number of environments. Genotypes with the smallest values of the superiority tend to be more stable, and closer to the best genotype in each environment.



Genotype	
AF12A-423-2/14	0.3350
AF12A-423-6/21	0.4751
AF12A-423-6/22	0.4941
AF12A-423-4/21	0.5315
AF12A-423-4/22	0.6395
AF12A-210-2/6	0.7153
AF12A-210-2/7	0.7450
AF12A-210-15/23	0.7543
AF12A-210-2/12	0.8175
AF12A-210-2/9	0.8245
AF12A-423-5/22	0.8866
AF12A-423-2/21	1.0259
Commercial Check 1	1.0582
Commercial Check 2	1.0649
AF12A-423-2/16	1.1475
AF12A-210-2/8	1.1640
AF12A-209-2/8	1.2075
AF12A-423-6/16	1.2515
AF12A-209-3/8	1.3652
AF12A-423-4/16	1.3668

Static Stability Measures Coefficients

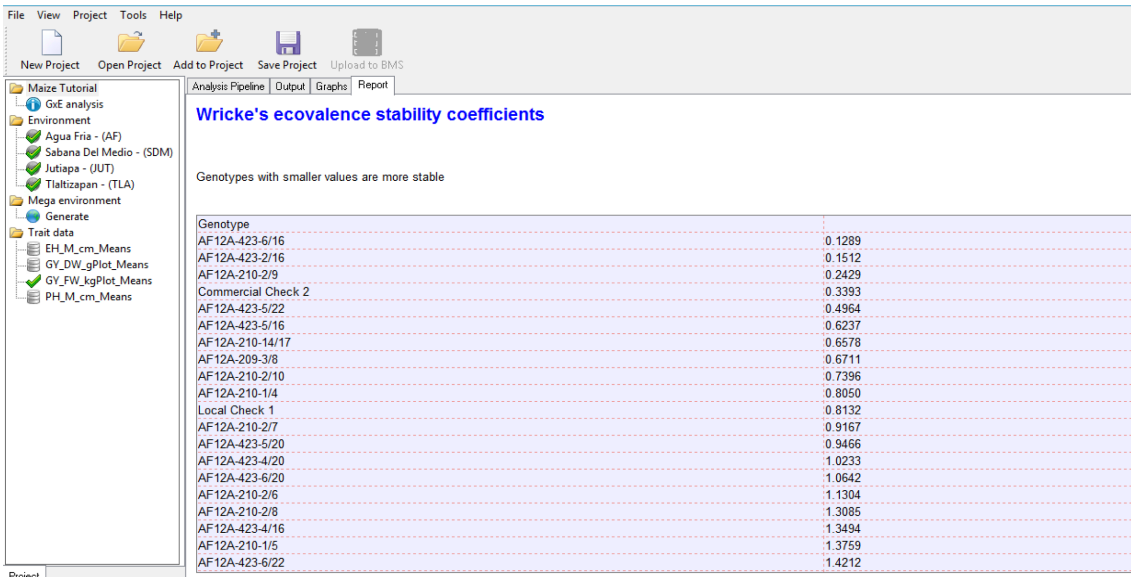
The Static Stability Coefficient is defined as the variance around the germplasm's phenotypic mean across all environments. This provides a measure of the consistency of the genotype, without accounting for performance.

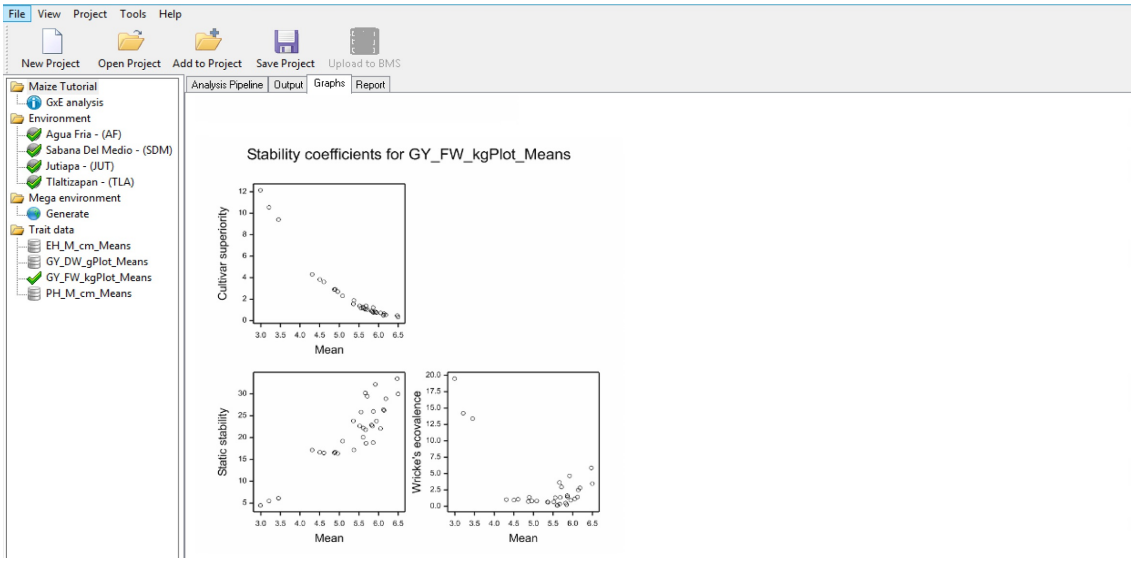


Genotype	
AF12A-423-4/23	4.45
AF12A-423-5/23	5.48
AF12A-423-2/23	6.11
AF12A-210-1/4	16.35
AF12A-423-6/20	16.46
AF12A-210-2/10	16.48
AF12A-423-5/20	16.61
AF12A-210-1/5	16.68
AF12A-423-4/20	17.13
AF12A-423-5/16	17.15
AF12A-423-4/16	18.69
AF12A-209-2/8	18.84
Local Check 1	19.17
AF12A-423-6/16	20.07
Commercial Check 2	21.81
AF12A-210-2/6	22.06
AF12A-423-2/16	22.19
AF12A-210-2/9	22.63
AF12A-209-3/8	22.64
AF12A-423-5/22	22.86

Wrick's Ecovalence Stability Coefficients

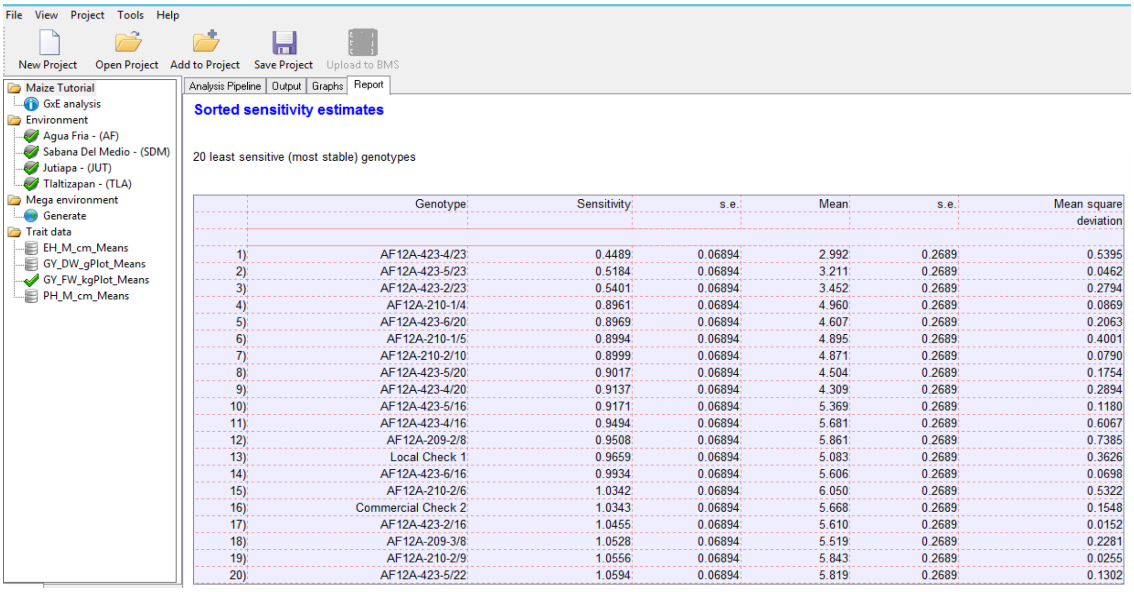
Wricke's Ecovalence Stability Coefficient (Wricke, 1962) is the contribution of each genotype to the genotype-by-environment sum of squares, in an un-weighted analysis of the genotype-by-environment means. A low value indicates that the genotype responds in a consistent manner to changes in environment; i.e. stable from a dynamic point of view.
Like static stability, the Wricke's Ecovalence does not account for genotype performance.



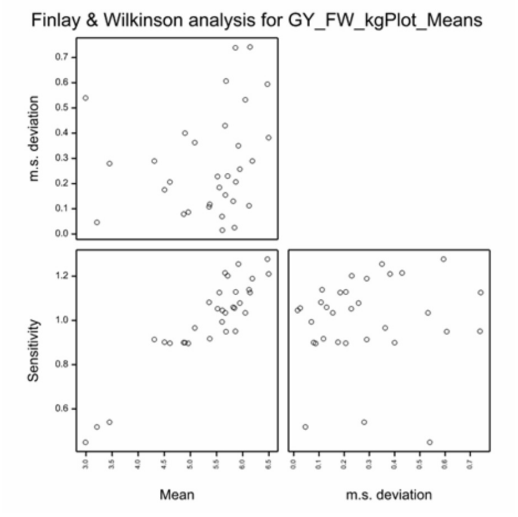


Finlay and Wilkinson Modified Joint Regression Analysis

The Finlay and Wilkinson Modified Joint Regression Analysis ranks germplasm based on phenotypic stability for each individual trait.

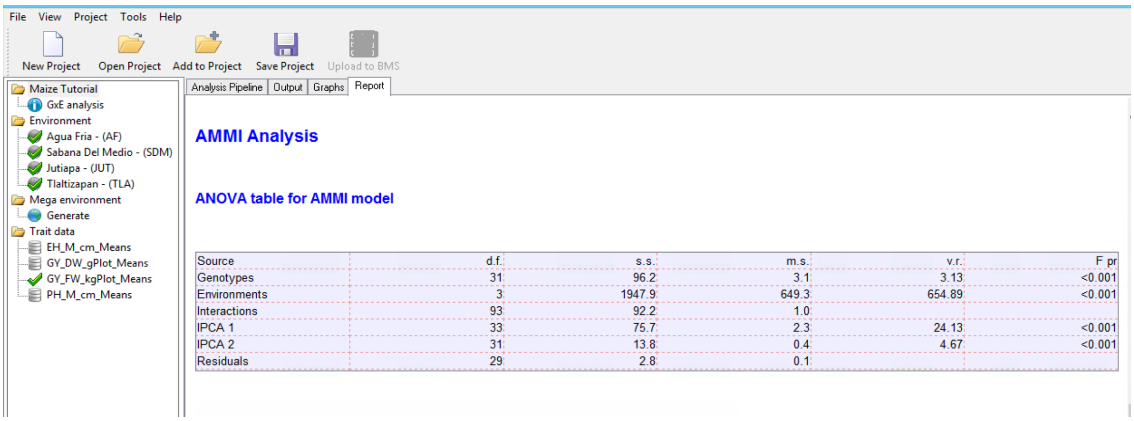


	Genotype	Sensitivity	s.e.	Mean	s.e.	Mean square deviation
1)	AF12A-423-4/23	0.4489	0.06894	2.992	0.2689	0.5395
2)	AF12A-423-5/23	0.5184	0.06894	3.211	0.2689	0.0462
3)	AF12A-423-2/23	0.5401	0.06894	3.452	0.2689	0.2794
4)	AF12A-210-1/4	0.8961	0.06894	4.960	0.2689	0.0869
5)	AF12A-423-6/20	0.8969	0.06894	4.607	0.2689	0.2063
6)	AF12A-210-1/5	0.8994	0.06894	4.895	0.2689	0.4001
7)	AF12A-210-2/10	0.8999	0.06894	4.871	0.2689	0.0790
8)	AF12A-423-5/20	0.9017	0.06894	4.504	0.2689	0.1754
9)	AF12A-423-4/20	0.9137	0.06894	4.309	0.2689	0.2894
10)	AF12A-423-5/16	0.9171	0.06894	5.369	0.2689	0.1180
11)	AF12A-423-4/16	0.9494	0.06894	5.681	0.2689	0.6067
12)	AF12A-209-2/8	0.9508	0.06894	5.861	0.2689	0.7385
13)	Local Check 1	0.9659	0.06894	5.083	0.2689	0.3626
14)	AF12A-423-6/16	0.9934	0.06894	5.606	0.2689	0.0698
15)	AF12A-210-2/6	1.0342	0.06894	6.050	0.2689	0.5322
16)	Commercial Check 2	1.0343	0.06894	5.668	0.2689	0.1548
17)	AF12A-423-2/16	1.0455	0.06894	5.610	0.2689	0.0152
18)	AF12A-209-3/8	1.0528	0.06894	5.519	0.2689	0.2281
19)	AF12A-210-2/9	1.0556	0.06894	5.843	0.2689	0.0255
20)	AF12A-423-5/22	1.0594	0.06894	5.819	0.2689	0.1302

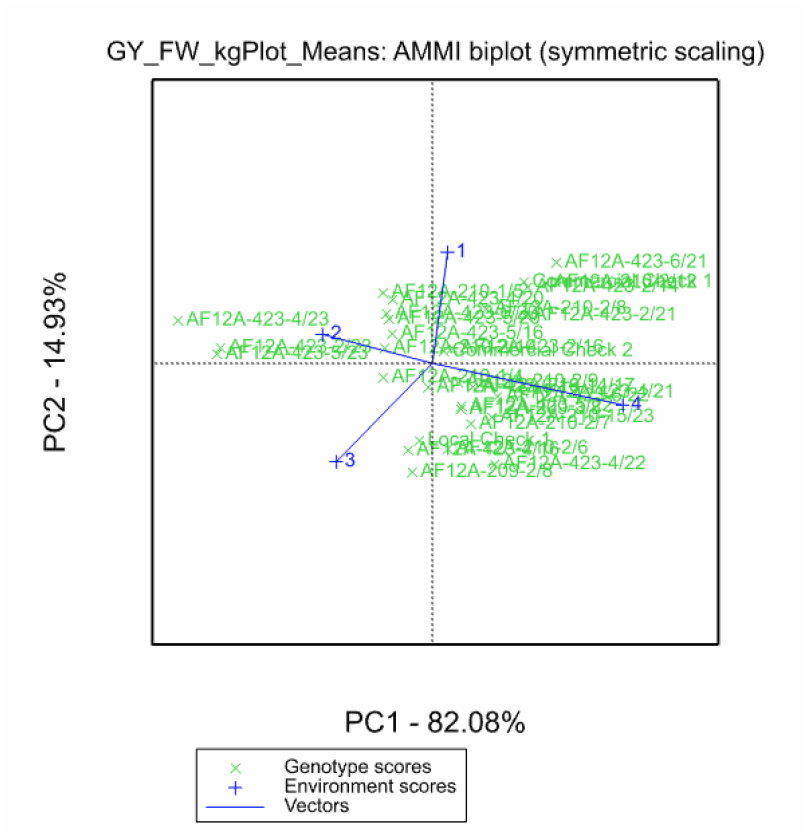


AMMI Model

In the Additive Main Effects and Multiplicative Interaction (AMMI) model, a two-way ANOVA additive model is performed (additive main effects), followed by a principal component analysis on the residuals (multiplicative interaction). As a result, the interaction is characterized by Interaction Principal Components (IPCA), where genotypes and environments can be simultaneously plotted in biplots.

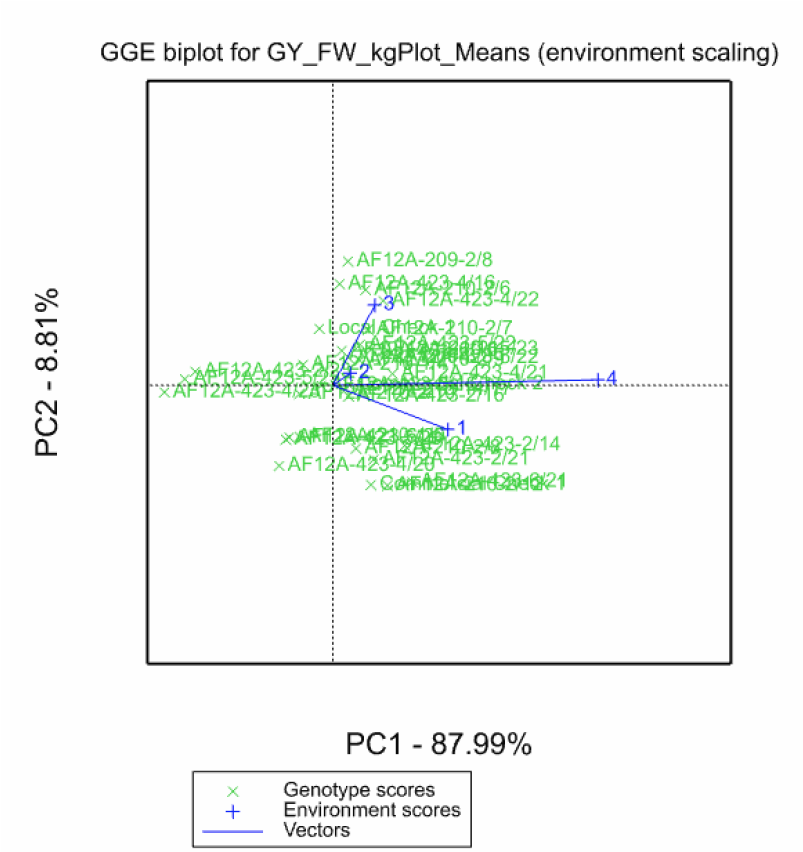


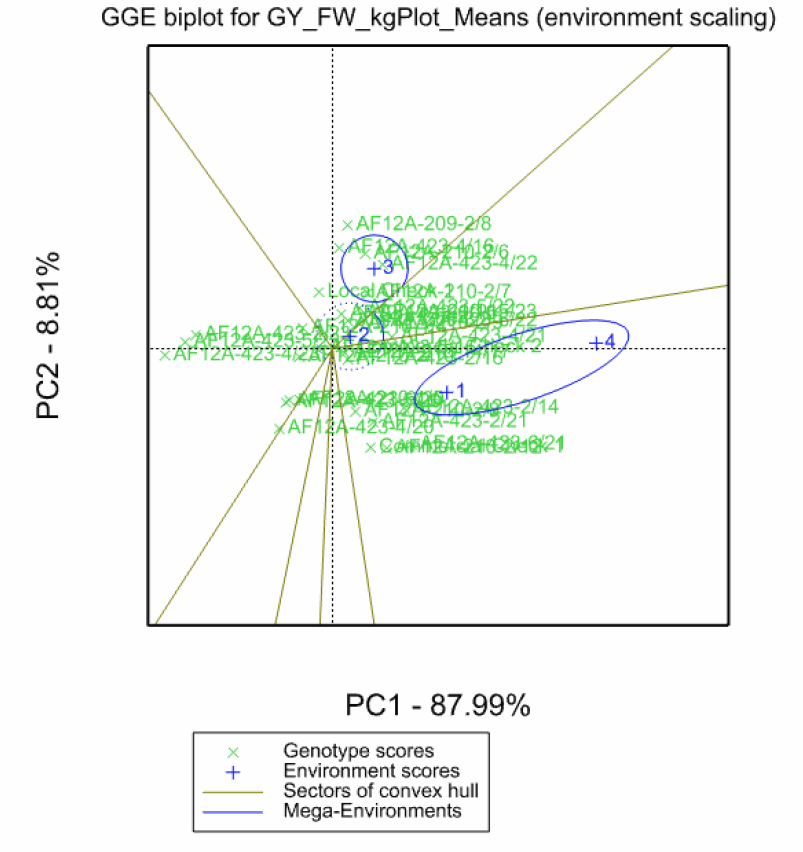
Source	d.f.	s.s.	m.s.	v.r.	F pr
Genotypes	31	96.2	3.1	3.13	<0.001
Environments	3	1947.9	649.3	654.89	<0.001
Interactions	93	92.2	1.0		
IPCA 1	33	75.7	2.3	24.13	<0.001
IPCA 2	31	13.8	0.4	4.67	<0.001
Residuals	29	2.8	0.1		



GGE Model

In the Genotype Main Effects and Genotype × Environment Interaction Effects (GGE) model (Yan et al. 2000 & 2003) a 1-way ANOVA, including environment as a main effect, is run followed by a principal component analysis on the residuals. Like AMMI, principal component scores can be used to construct biplots. Unlike the AMMI Model, in GGE the genotypic main effects are also represented in the plot. The GGE model is superior to AMMI analysis at differentiating mega-environments (Yan et al. 2007)

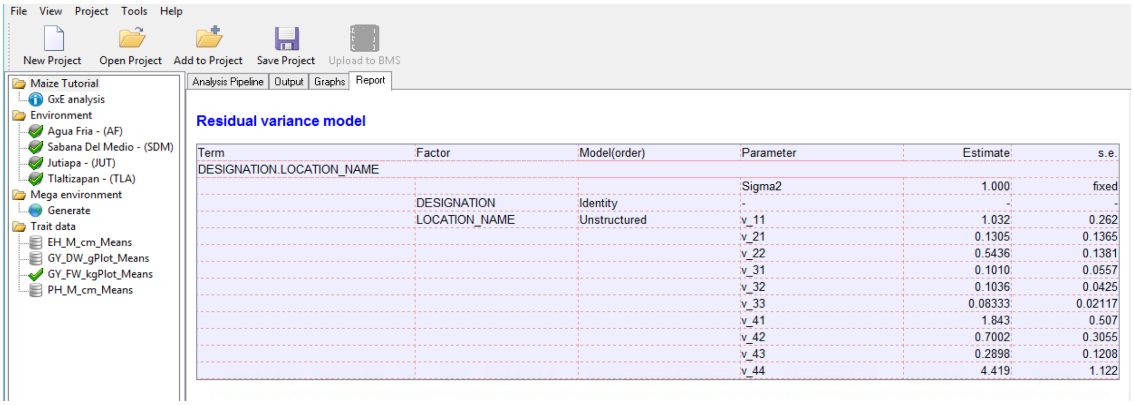




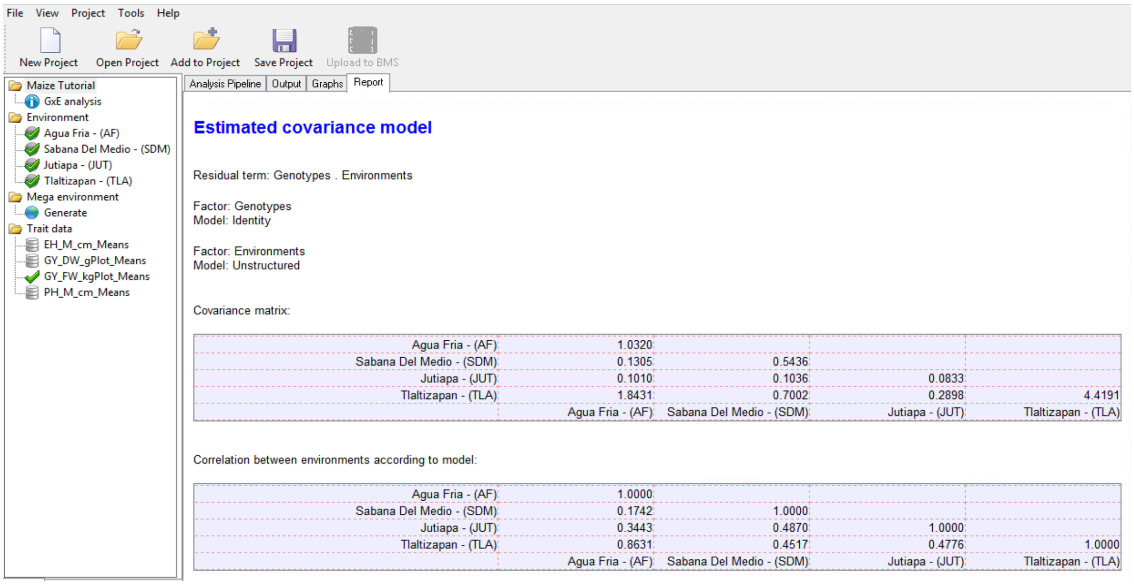
Environments 1 & 4 cluster, indicating that these two locations have similar environmental effects on phenotype and small GxE interactions.

Variance-Covariance Model & Correlation Matrix

Details on the variance-covariance model, including the pairwise correlation matrix from the covariance model is presented in a table in the Report tab. In the correlation matrix values close to 1 indicate higher correlation between environments. A value of 1 indicates a perfect correlation, such as when an environment is compared to itself.



Term	Factor	Model(order)	Parameter	Estimate	s.e.
DESIGNATION.LOCATION_NAME					
			Sigma2	1.000	fixed
	DESIGNATION	Identity	-	-	-
	LOCATION_NAME	Unstructured	v_11	1.032	0.262
			v_21	0.1305	0.1365
			v_22	0.5436	0.1381
			v_31	0.1010	0.0557
			v_32	0.1036	0.0425
			v_33	0.08333	0.02117
			v_41	1.843	0.507
			v_42	0.7002	0.3055
			v_43	0.2898	0.1208
			v_44	4.419	1.122



Agua Fria - (AF)	1.0320			
Sabana Del Medio - (SDM)	0.1305	0.5436		
Jutiapa - (JUT)	0.1010	0.1036	0.0833	
Tlaltizapan - (TLA)	1.8431	0.7002	0.2898	4.4191
	Agua Fria - (AF)	Sabana Del Medio - (SDM)	Jutiapa - (JUT)	Tlaltizapan - (TLA)

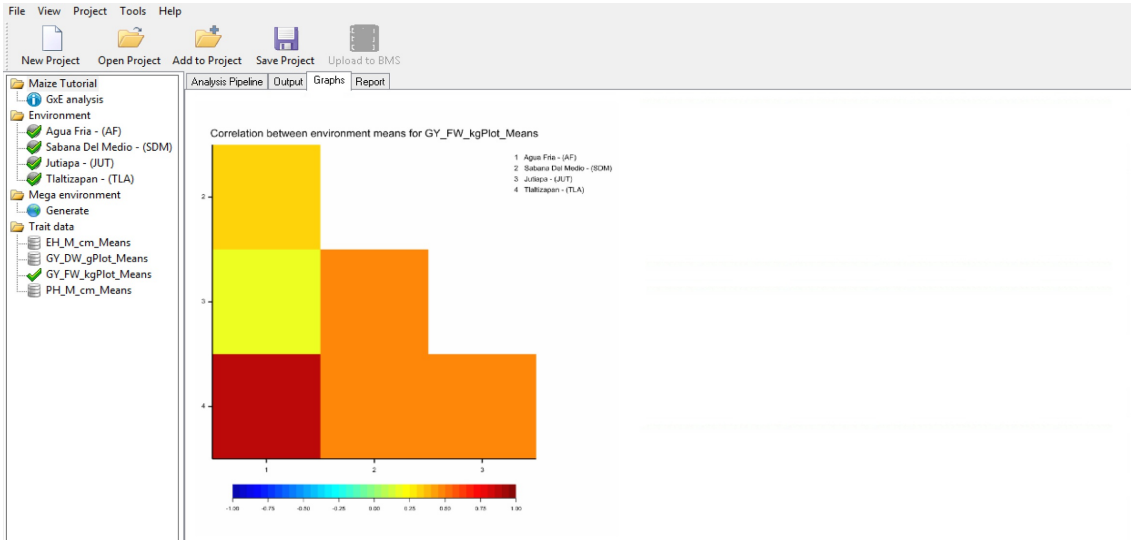
Correlation between environments according to model:

Agua Fria - (AF)	1.0000			
Sabana Del Medio - (SDM)	0.1742	1.0000		
Jutiapa - (JUT)	0.3443	0.4870	1.0000	
Tlaltizapan - (TLA)	0.8631	0.4517	0.4776	1.0000
	Agua Fria - (AF)	Sabana Del Medio - (SDM)	Jutiapa - (JUT)	Tlaltizapan - (TLA)

*Correlation Matrix for Grain Yield (*GY_FW_kgPlot*): Environment 1 is most positively correlated to the Environment 4 (0.8631), suggesting that the two locations have similar environmental effects on phenotype.*

Correlation Heat Map

The correlation heat matrix visualizes correlations with color; warm colors (red) indicating high positive correlation between environments, and cool colors (blue) indicating high negative correlation between environments.

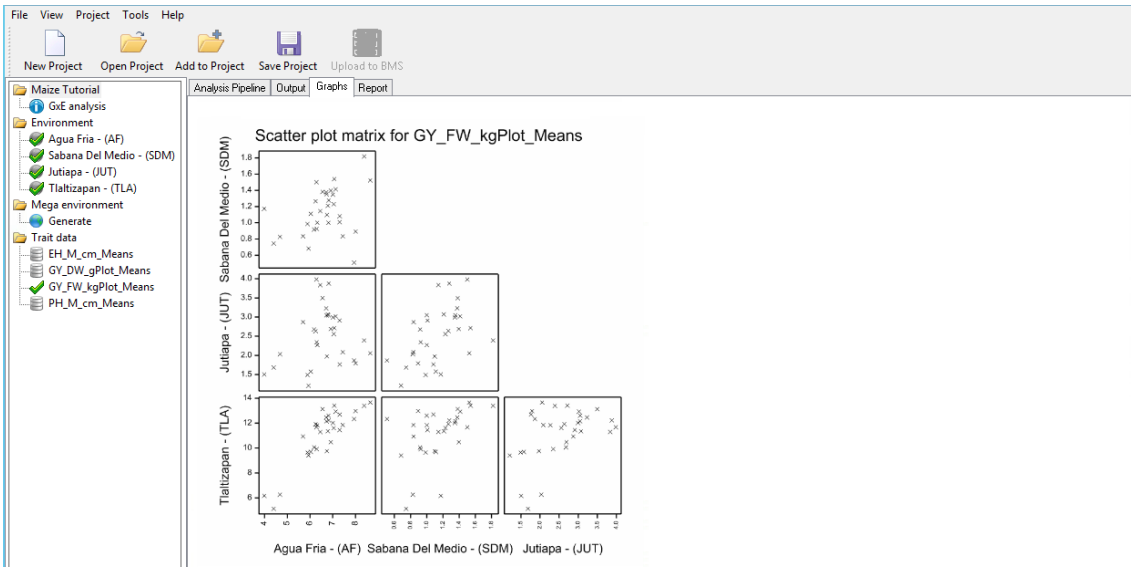


*Correlation Heat Map of Grain Yield (*GY_FW_kgPlot*): Environment 1 is most positively correlated (red) to Environment 4*

environment, suggesting that these two locations have similar environmental effects on phenotype and small GxE interactions.

Scatter Plot Matrix

The scatter plot matrix illustrates the association of genotypic performance between each pair of environments.



*Scatter Plot Matrix for Grain Yield (*GY_FW_kgPlot*): A positive correlation is observed between genotypic performance at Environments 1 & 4 indicating similar environmental effects on phenotype for these environments and small GxE interactions. However, little correlation is observed between genotypic performance at Environment 1 and 3, indicating large GxE interactions between these two environments.*

References

Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. Biometrics, 44, 705–715.

Gauch, H.G. (1992). Statistical Analysis of Regional Yield Trials – AMMI analysis of factorial designs. Elsevier, Amsterdam.

Finlay, K.W. & Wilkinson, G.N. (1963). The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research, 14, 742-754.

Murray, D. Payne, R, & Zhang, Z. (2014) Breeding View, a Visual Tool for Running Analytical Pipelines: User Guide. VSN International Ltd. (.pdf) (Sample data .zip).

Lin, C.S. & Binns. M.R. (1988). A superiority performance measure of cultivar performance for cultivar x location data. Canadian Journal of Plant Science, 68, 193-198.

Lin, C. S., Binns, M. R., & Lefkovitch, L. P. (1986). Stability analysis: Where do we stand? Crop Science, 26, 894–900.

Oakey, H., Verbyla, A. P., Pitchford, W., Cullis, B., & Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. Theoretical and Applied Genetics, 113, 809–819.

Wricke, G. (1962). Uber eine method
zur erfassung der okogischen streubreite in feldversuchen. Zeitschriff Fur Pflanzenzuchtung, 47, 92-96.

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Science, 40, 597–605.

Yan, W. & Kang, M.S. (2003). GGE Biplot Analysis: a Graphical Tool for Breeders, Geneticists and Agronomists. CRC Press, Boca Raton.

Yan, W., Kang, M.S. Ma, B., Woods, S., Cornelius, P.L. (2007) GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data. Crop Science. 47, 643–653.