DOI: 10.5433/1679-0359.2020v41n3p767

Key locations for soybean genotype assessment in South Brazil region

Locais-chave para a avaliação de genótipos de soja na região Sul do Brasil

Samuel Cristian Dalló[1]; Andrei Daniel Zdziarski[1]; Leomar Guilherme Woyann[2]; Anderson Simionato Milioli[3]; Rodrigo Zanella[1]; Vinícius de Bitencourt Bez Batti[1]; Giovani Benin[4*]

Highlights:

Representative and discriminating locations to select genotypes were identified.
Non-representative and redundant locations were identified and can be excluded.
The best locations for evaluating VCU trials and to reduce costs were recommended.

Abstract

Mitigating the high costs of soybean breeding programs requires constant improvement of all the involved processes. Identifying representative and discriminating test locations, as well as excluding redundant and/or non-representative locations, makes it possible to select genotypes with more accuracy while reducing the costs of the multi-environment trials (MET). Therefore, this study had three objectives: to evaluate the representativeness and discriminating power of test locations, to identify similar test locations for each Edaphoclimatic Region (ECR) and locations that did not contribute to genotype evaluation, and to recommend the best locations for evaluating MET to reduce breeding program costs in the soybean macro-regions 1 (M1) and 2 (M2). Grain yield (GY) data from 'Value-for-Cultivation-and-Use' (VCU) trials obtained during the 2012-2016 crop seasons were used, totaling 132 environments (location x year) and 43 genotypes. The experiments were arranged in a randomized complete block design with three replications. Representative and discriminant locations were identified by GGL (genotype main effects plus genotype × location interaction) + GGE (genotype main effects plus genotype × environment interaction) analysis, using GGEbiplot software. Representative and discriminant locations were identified for each ECR and can be used as core locations for breeding programs. Similarly, locations that were not representative and discriminant, or that present redundancy in the results, should be excluded from or replaced in MET. The most recommended locations for conducting VCU trials in M1 were Cachoeira do Sul (ECR 101); Ronda Alta, Passo Fundo, Santa Bárbara do Sul, and Ciríaco (ECR 102); and Castro (ECR 103). For M2, the most suitable locations are Rolândia, Marechal Cândido Rondon, Campo Mourão, Santa Terezinha de Itaipu, Palotina, Floresta, and Londrina (ECR 201); Naviraí (ECR 202); and Ponta Porã and Maracajú (ECR 204).

Key words: *Glycine max* (L.) Merrill. GGL + GGE Biplots. Representativeness and Discriminating power.

---

[1] Mestres em Agronomia, Programa de Pós-Graduação em Agronomia, Universidade Tecnológica Federal do Paraná, UTFPR, Pato Branco, PR, Brasil. E-mail: samueldallo@hotmail.com, dz_andrei@hotmail.com, zanella_rodrigo@hotmail.com, viniciusbezbatti95@gmail.com

[2] Pós-Doutorado, Programa de Pós-Graduação em Agronomia, UTFPR, Pato Branco, PR, Brasil. E-mail: leowoyann@gmail.com

[3] Discente do Curso de Doutorado do Programa de Pós-Graduação em Agronomia, UTFPR, Pato Branco, PR, Brasil. E-mail: milioli.utfpr@gmail.com

[4] Prof. Dr., Programa de Pós-Graduação em Agronomia, UTFPR, Pato Branco, PR, Brasil. E-mail: benin@utfpr.edu.br

[*] Author for correspondence

Received: Apr. 03, 2018 - Approved: Mar. 25, 2020

Resumo

A redução dos elevados custos dos programas de melhoramento de soja requer melhoria constante de todos os processos envolvidos. A identificação de locais de teste representativos e discriminativos, bem como a exclusão de locais redundantes e/ou não representativos, possibilita a seleção de genótipos com maior precisão, e ao mesmo tempo reduz os custos dos ensaios multiambientes (MET). Portanto, este estudo teve três objetivos: avaliar a representatividade e a discriminância dos locais de teste; identificar locais de teste semelhantes e que não contribuíram para a avaliação genotípica dentro de cada Região Edafoclimática (ECR); e recomendar os melhores locais para a avaliação de MET, a fim de reduzir os custos do programa de melhoramento genético nas macrorregiões 1 (M1) e 2 (M2) de adaptação da soja. Dados de rendimento de grãos (RG) de ensaios de "Valor de Cultivo e Uso" (VCU) obtidos durante os anos agrícolas 2012-2016 foram utilizados, totalizando 132 ambientes (local x ano) e 43 genótipos. Os experimentos foram conduzidos em delineamento de blocos completamente casualizados, com três repetições. Locais representativos e discriminantes foram identificados pela análise GGL (efeito principal de genótipo + interação genótipo × local) + GGE (efeito principal de genótipo + interação genótipo × ambiente), utilizando o software GGEbiplot. Locais representativos e discriminantes foram identificados para cada ECR, e podem ser utilizados como locais-chave em programas de melhoramento. Da mesma forma, locais que não foram representativos e discriminantes, ou que apresentam redundância foram identificados e devem ser excluídos ou substituídos no MET. Os locais recomendados para a condução de ensaios de VCU na M1 são: Cachoeira do Sul (ECR 101); Ronda Alta, Passo Fundo, Santa Bárbara do Sul e Ciríaco (ECR 102); e Castro (ECR 103). Já para M2, os locais recomendados são: Rolândia, Marechal Cândido Rondon, Campo Mourão, Santa Terezinha de Itaipu, Palotina, Floresta e Londrina (ECR 201); Naviraí (ECR 202); e Ponta Porã e Maracaju (ECR 204).

Palavras-chave: *Glycine max* (L.) Merrill. GGL + GGE biplots. Representatividade e discriminância.

Introduction

Soybean is cultivated over a wide area across different geographic regions of Brazil, which show significant variability in soil type, temperature, rainfall, and photoperiod. Under these differing conditions, soybean genotype performance is strongly influenced by genotype × environment interaction (GEI), making it difficult to select and recommend adapted and stable genotypes across wide environmental variations.

Before the release of a new cultivar, genotypes (lineages) are evaluated and selected in multi-environment trials (MET). In Brazil, the final evaluation is performed in METs called 'Value-for-Cultivation-and-Use' (VCU) trials. These trials are evaluated in edaphoclimatic regions (ECRs) that have previously been defined by the Brazilian Ministry of Agriculture, Livestock and Supply and should be conducted in each ECR of interest for at least two years before the release of a new cultivar.

An extensive MET network is required to obtain accurate information on the performance of genotypes in ECRs. However, some test locations may present similar biological characteristics, showing a similar variability in genotype performance (Luo et al., 2015; Yan, 2015) and getting redundant results. Bearing in mind the high costs of conducting these trials, the correct choice of test locations is crucial (Yan, Pageau, Frégeau-Reid, & Durand, 2011). Therefore, it is important to eliminate or replace locations that have similar test result patterns to improve selection efficiency.

An ideal test location must present a high discriminative capacity to identify the differences between genotypes (Dia et al., 2016). In addition, a test location should also be representative of the other locations in the target region (TR), so that the selected genotypes perform similarly across sites (Khatun, Islam, Anisuzzaman, Ahmed, & Haque, 2015; Luo et al., 2015).

Biplot analyses have been widely used by many authors in evaluating the best test locations, as shown by Luo et al. (2015), Yan (2015, 2016), and Dia et al. (2016). In GGE (genotype main effects plus genotype × environment interaction) biplot analyses, the effect of years is evaluated in an individualized way. Thus, it is necessary to compare the results year by year; which may, however, lead to uncertainties in the results obtained because of the pronounced effect between the evaluation years, indicating that it may not be possible to identify repetitive patterns over the years (Yan, 2014). To visualize patterns in genotype responses for different environments (locations × years), a GGL + GGE [(genotype main effects plus genotype × location interaction) + (genotype main effects plus genotype × environment interaction)] analysis is considered superior to year-by-year analyses, since it summarizes patterns in a single biplot (Yan, 2014, 2015). This analysis allows G + GE and G + GL patterns to be visualized, in addition to any repeatable GL in the dataset (Yan, 2014, 2015, 2016).

Therefore, this study had three objectives: to evaluate the representativeness and discriminating power of test locations, to identify similar test locations for each ECR and locations that did not contribute to genotype evaluation, and to recommend the best locations for evaluating VCU trials to reduce breeding program costs.

Materials and Methods

Data source and experimental methods

Grain yield (GY) data from VCU trials conducted in the crop seasons of 2012/13, 2013/14, 2014/15, 2015/16, and 2016/17 were used. The trials were conducted in ECRs 101, 102 and 103 in macro-region 1 (M1), and in ECRs 201, 202, and 204 in macro-region 2 (M2). Altogether, 34 locations and 43 genotypes were evaluated (Supplementary Tables 1 and 2, and Figure 1). Each ECR was considered as a mega-environment (ME), defined according to the third regional approximation of the VCU trials (Kaster & Farias, 2012), which is the TR under study. The trials were conducted using a randomized complete block design (RCBD) with three replications. The experimental units consisted of four rows, each 5 m in length, with 0.5 m spacing between rows. The useful area of the plot consisted of the two central lines, totaling an area of 5 m^2. The crop was harvested at maturity and the GY per plot (kg $plot^{-1}$) was evaluated. Subsequently, moisture was corrected to 13%, and the GY was calculated as kg ha^{-1}. Cultivation practices were carried out following the technical recommendations for soybean.

Table 1

Numerical values for environment projections on the average environment axis (AEA; AEC_X), and their distances to the AEA (AEC_Y), and numerical values for the discriminating ability (vector length) and representativeness (correlation with the AEA) of the environments for edaphoclimatic region (ECR) 101

Environments*	AEC_X	AEC_Y	Vector Length	Correlation with AEA
CACS-L_12	1.45	0.21	1.46	0.99
CACS-R_13	1.41	-0.1	1.41	1
CACS-L_13	1.21	-0.09	1.21	1
CACS-R_15	1.17	-0.2	1.19	0.99
CACS-L_14	1.09	0.83	1.37	0.8
CACS-R_14	1.06	-0.69	1.27	0.84
SBRJ_12	0.97	-0.71	1.2	0.81

continue

continuation

SBRJ_14	0.92	-0.24	0.95	0.97
SBRJ_13	0.33	0.26	0.42	0.79
CACS-R_12	-0.43	0.74	0.86	-0.5

*An environment is composed of a location and a year (for example CACS-L_12 = Cachoeira do Sul – Várzea, in the year 2012). The locations are Cachoeira do Sul – Várzea (CACS-L), Cachoeira do Sul – Sequeiro (CACS-R), and São Borja (SBRJ).

Table 2
Numerical values for environment projections on the average environment axis (AEA; AEC_X) and their distances to the AEA (AEC_Y), and numerical values for the discriminating ability (vector length) and representativeness of the environments for edaphoclimatic region (ECR) 102

Environments*	AEC_X	AEC_Y	Vector Length	Correlation with AEA
STBS_12	1.74	0.11	1.74	1
STBS_15	1.69	0.28	1.71	0.99
PSFD_13	1.54	-0.2	1.55	0.99
STBS_13	1.46	-0.06	1.46	1
STBS_14	1.46	0.89	1.71	0.86
PSFD_16	1.46	-0.64	1.59	0.92
PSFD_14	1.42	-0.27	1.45	0.98
STBS_16	1.39	-0.11	1.39	1
CMPV_13	1.34	-0.18	1.35	0.99
RALT_12	1.3	-0.46	1.38	0.94
RALT_15	1.3	-0.11	1.3	1
CRCO_14	1.24	-0.51	1.34	0.93
PSFD_12	1.21	-0.34	1.25	0.96
CNDR_15	1.14	-0.38	1.2	0.95
CNDR_14	1.01	-0.76	1.26	0.8
PSFD_15	1.01	0.69	1.22	0.83
SLGO_14	0.98	1.61	1.88	0.52
CMPV_12	0.83	0.01	0.83	1
TPJR_13	0.75	-0.48	0.89	0.84
CMPV_15	0.7	0.54	0.88	0.79
CRCO_15	0.64	0.61	0.89	0.73
TPJR_15	0.51	0.75	0.91	0.56
CNDR_16	0.3	-0.65	0.72	0.41
SLGO_15	0.21	-0.06	0.22	0.96
SLGO_12	-0.06	1.29	1.29	-0.05
CMPV_16	-0.14	-0.68	0.69	-0.21
CMPV_14	-0.26	-0.48	0.55	-0.47
TPJR_16	-0.31	-0.39	0.5	-0.62

*An environment is composed of a location and a year (for example STBS_12 = Santa Bárbara do Sul, in the year 2012). The locations are Santa Bárbara do Sul (STBS), Passo Fundo (PSFD), Campos Novos (CMPV), Ronda Alta (RALT), Ciríaco (CRCO), Condor (CNDR), São Luiz Gonzaga (SLGO), and Tapejara (TPJR).

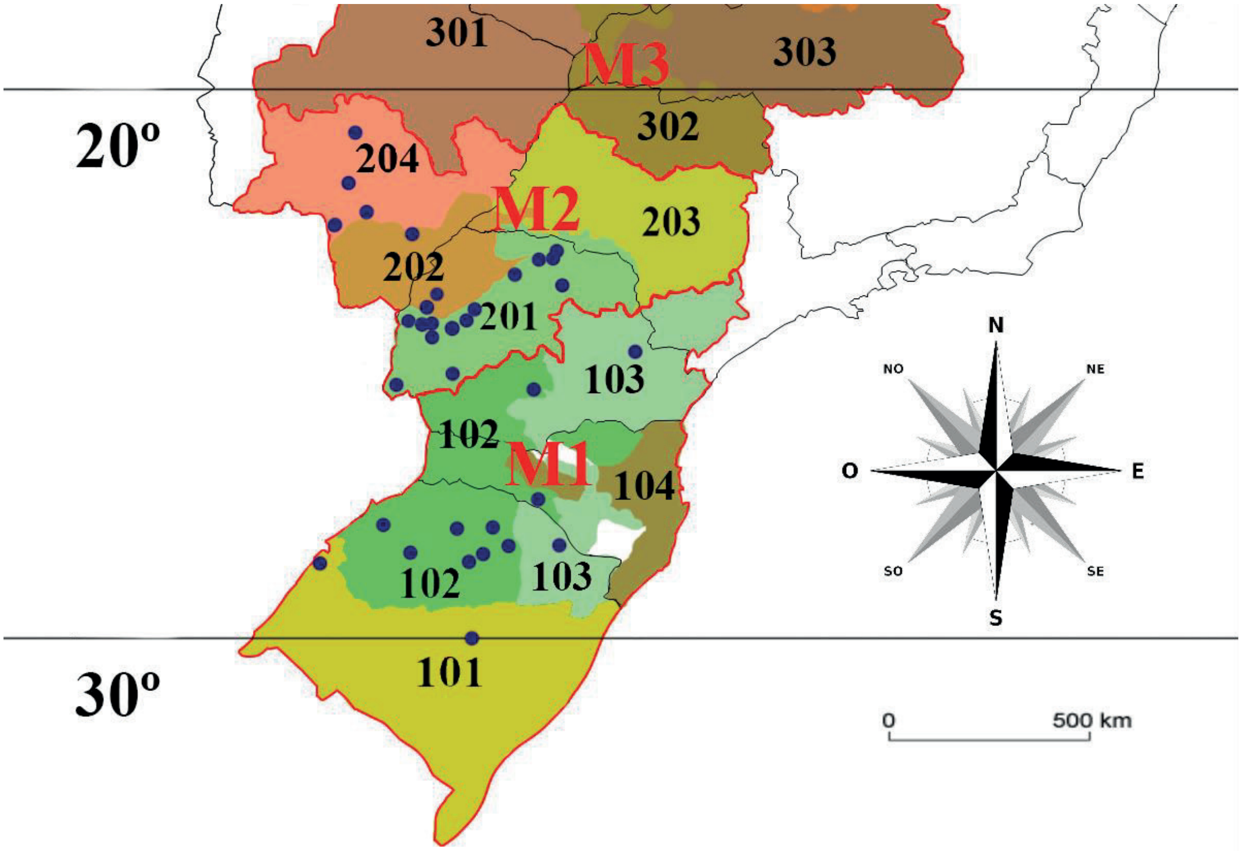

Figure 1. Test locations used in analyses to identify the best location for soybean Value-for-Cultivation-and-Use (VCU) trials in Brazil. Blue dots represent the locations.

Statistical methods

The GGL + GGE analysis was performed using GGEbiplot software (Yan, 2001); this analysis allows the most representative locations in a TR to be identified. The cosine of the angle between the environmental vector and average environment is the approximate genetic correlation (r_g) between the environments (Yan, 2014); the length of the vector indicates the consistency of the results between the years of testing. In addition, when the biplot explanation is high and the scale is based on standard deviation (SD) and adjusted heritability (*h*; scaling 2), the vector length approximates to the square root of *h* (Yan, 2014), i.e. the discrimination power. In this sense, representativeness is considered more important than discrimination in an integrated assessment, since representativeness varies between -1 and +1 (r_g) and discrimination (*h*) varies between 0 and 1, with common values between 0.5 and 1 (Yan, 2014). An ideal environment would therefore have values of $r_g = 1$ and $h = 1$, in addition to an integrated index ($r_g h$) = 1 that is representative of the ME in which it is inserted, as well as having the capacity to discriminate superior genotypes efficiently (Yan, 2014). On the other hand, locations with a short vector are not desirable, because they are poorly associated with, and/or have a lower heritability than other evaluated locations when the fit is less than perfect (Yan, 2014).

Analysis of G vs. GE (genotypic effect vs. genotype × environment effects) allows the ability of environments to select superior genotypes to be inferred. In this analysis, the combination of location + year forms an environment that is plotted on the biplot graph. In this sense, biplots for selecting G vs. GE are divided into two components. The projection of the environments in the single-arrow average environment axis (AEA) line (red line) indicates the ability of the environment to select genotypes based on the G effect, where the further to the right side of the biplot, the more useful the environment is in selecting better genotypes (Yan, 2014). In addition, the smaller the vector length in the average environment coordinate (AEC) direction (double-arrow line, blue), the greater the ability of the location to select genotypes with the G effect, while the longer the vector length, the greater the ability to select by GE effect (Yan, 2014). Thus, environments with short vectors and positioned on the right-side of the biplot are more desirable, because they have a greater capacity to select superior genotypes.

For the GGE biplot analyses, the following parameters were used: Transform = 0 (without transformation), Scaling = 2 (SD scaled + h weighted), Centering = 2 (biplot based on G + GE), and singular value partitioning (SVP) = 2 (focus on environment). In the scale based on adjusted heritability (HA-GGE), the vector length approximates the square root of heritability and the angle between the environmental vectors approximates the r_g value between environments (Yan & Holland, 2010; Yan, 2014). The most appropriate scale for evaluating environments is HA-GGE because it graphically shows the relative utility of the environments in terms of response to selection (Yan & Holland, 2010).

Results

Regarding the GGL + GGE analyses for M1, the location CACS-R (Figure 2A) was revealed as the most representative for ECR 101 across all years. However, the most consistent results were observed for CACS-L, since it had the longest vector. In the analysis of G vs. GE (Figure 2D), it was observed that environments CACS-L_12 and CACS-R_13 presented the greatest capacity to select by the G effect, given their short vectors and position on the right side of the biplot. Environment CACS-R_12 was the only one positioned to the left side of the AEC line, and the vector projection relative to AEA (AEC_Y) was the second highest (0.74) (Table 1). Both this environment and SBRJ_13 presented the smallest capacities for selecting G because they are positioned on the left side of the biplot. In this sense, the cosine of the angle of SBRJ in relation to AEA indicated that this environment is not representative of ECR 101.

For ECR 102 (Figure 2B), the highest r_g values were observed between environments TPJR, CMPV, CRCO, RALT, PSFD, and STBS. On the other hand, environments CNDR and SLGO were not representative of this ECR, since they presented the highest angles in relation to AEA and, therefore, a lower r_g with the average environment. Further, CMPV and TPJR environments presented inconsistent results, since they had short vectors. In the G vs. GE analysis (Figure 2E, Table 2), it was also observed that environment STBS_12 had the greatest capacity to select genotypes using G effect since it was positioned on the right side of the biplot and had a short vector. On the other hand, environments TPJR_16, CMPV_14, CMPV_16, and SLGO_12 were positioned to the left side of the AEC line, indicating that they were inefficient in selecting superior genotypes. In addition, in the G vs. GE analysis, SLGO_12 and SLGO_14 presented the highest vectors among the evaluated environments, revealing that they would be inefficient in selecting genotypes by the G effect (Figure 2E).

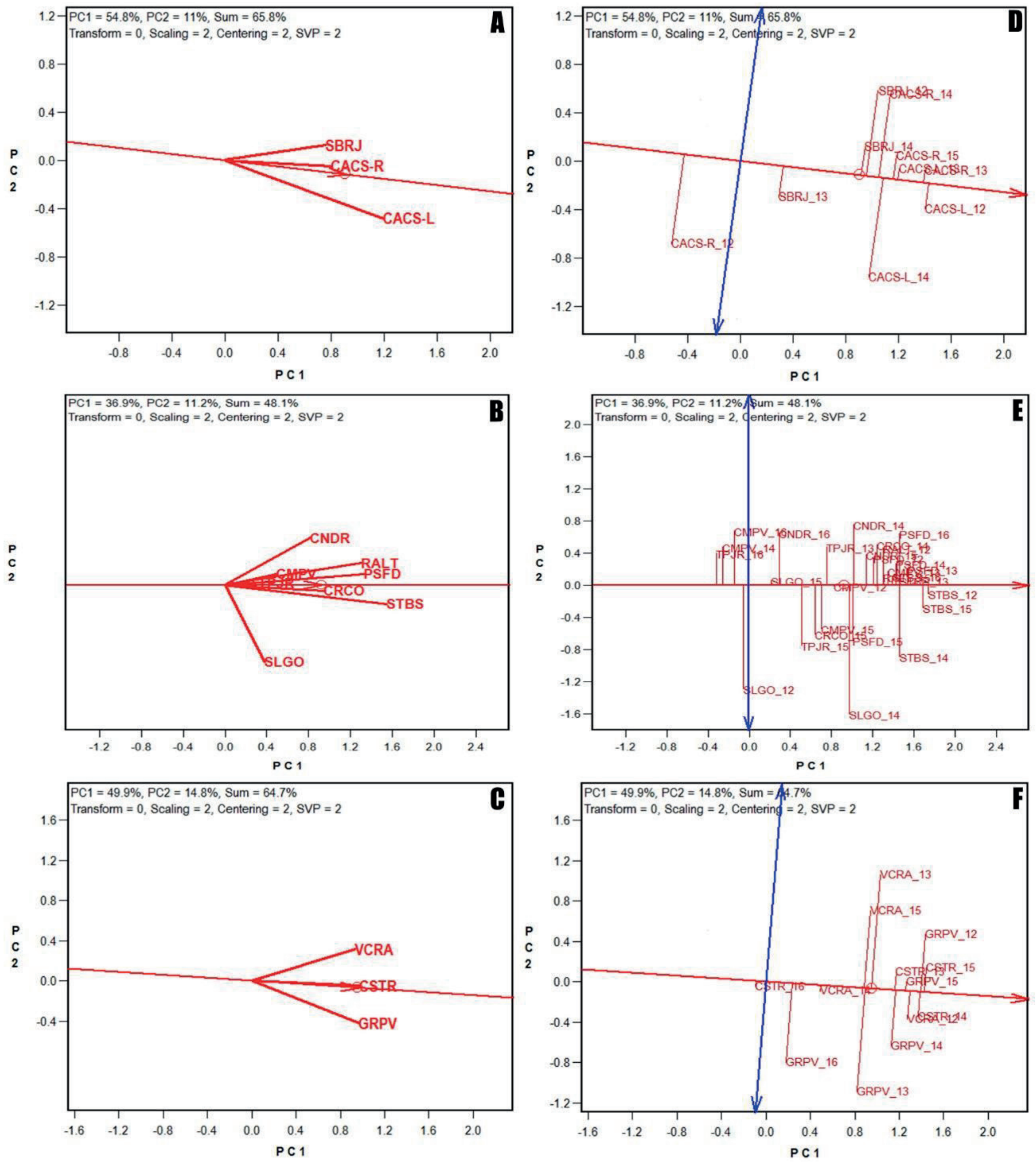

Figure 2. The GGL (genotype main effects plus genotype × location interaction) + GGE (genotype main effects plus genotype × environment interaction) biplot for macro-region 1 based on soybean crop data from 2012 to 2016 (101-A, 102-B, 103-C). The biplot form displays the ability of the environment to select soybean genotypes for G vs. GE (genotypic effect vs. genotype × environment effects) for macro-region 1 (101-D, 102-E, 103-F). A Tester is composed of a location and a year (for example SLGO_15 = São Luiz Gonzaga, in the year 2015). The locations are: Cachoeira do Sul – Lowland (CACS-L), Cachoeira do Sul – Rainfed (CACS-R), São Borja (SBRJ), Santa Bárbara do Sul (STBS), Passo Fundo (PSFD), Campos Novos (CMPV), Ronda Alta (RALT), Ciríaco (CRCO), Condor (CNDR), São Luiz Gonzaga (SLGO), Tapejara (TPJR), Castro (CSTR), Guarapuava (GRPV), and Vacaria (VCRA).

Regarding ECR 103 (Figure 2C), the mean of the five-year test showed location CSTR to be the most representative. In the G vs. GE analysis (Figure 2F), this environment also presented short vectors in all years, being the environment with the greatest capacity to select superior genotypes in 2015. However, in 2016, the value of environment CSTR_16 in relation to AEC_X was more negative in relation to AEC (-0.09), although only by a small magnitude (Table 3). Environments VCRA_13 and GRPV_13 presented longer vectors in the analysis of G vs. GE, respectively (Figure 2F).

Table 3
Numerical values for environment projections on the average environment axis (AEA; AEC_X) and their distances to the AEA (AEC_Y), and numerical values for the discriminating ability (vector length) and representativeness of the environments for edaphoclimatic region (ECR) 103

Environments*	AEC_X	AEC_Y	Vector Length	Correlation with AEA
CSTR_15	1.43	-0.25	1.45	0.99
GRPV_12	1.4	-0.58	1.51	0.92
CSTR_14	1.39	0.24	1.41	0.99
VCRA_12	1.3	0.27	1.33	0.98
GRPV_15	1.26	-0.09	1.27	1
GRPV_14	1.17	0.55	1.3	0.91
CSTR_13	1.16	-0.18	1.18	0.99
VCRA_13	0.95	-1.13	1.48	0.64
GRPV_13	0.9	1.03	1.37	0.66
VCRA_15	0.89	-0.77	1.18	0.76
VCRA_14	0.5	0.06	0.5	0.99
GRPV_16	0.24	0.79	0.82	0.29
CSTR_16	-0.09	0.06	0.1	-0.84

*An environment is composed of a location and a year (for example CSTR_15 = Castro, in the year 2015). The locations are Castro (CSTR), Guarapuava (GRPV), and Vacaria (VCRA).

In relation to ECRs, in M2, ECR 201 presented 54 environments (location + year) (Figures 3A and 3D). Thus, the large number of locations plotted on the graph makes the interpretation of the G vs. GE analysis difficult. In this sense, the evaluation of the values in the tables is fundamental for a better understanding of the patterns (Tables 4, 5, and 6). However, visual analysis remains useful, specifically for identifying sites to the left-side of the AEC line, which are undesirable, because of their low ability to select superior genotypes. Locations MRDN, PLTN, FLRT, STIU, MAMB, CMMR, RLDA, and LDRN presented the highest r_g, with the average environment in ECR 201 and the smallest angles between the environment vectors also indicated a high r_g between them (Figure 3A). However, in the G vs. GE analysis (Figure 3D), MAMB_13 and RLDA_13 were positioned to the left side of the AEC line and were, therefore, shown to be inefficient in selecting superior genotypes. Environments CCVL_13, CCVL_15, CCVL_16, CFLD_16, and TOLD_16 were also positioned left-side of the AEC line. In these environments, the value of AEC_X was negative; thus, these environments are inefficient in selecting superior genotypes (Figure 3D, Table 4). It was observed that the vector length of the environment STIU in the G vs. GE analysis presented average values, indicating that this location was efficient in selecting genotypes by G effect.

Table 4
Numerical values for the environment projections on the average environment axis (AEA; AEC_X) and their distances to the AEA (AEC_Y), and numerical values for the discriminating ability (vector length) and representativeness of the environments for edaphoclimatic region (ECR) 201

Environments*	AEC_X	AEC_Y	Vector Length	Correlation with AEA
RLDA_15	1.44	-0.44	1.5	0.96
MRDN_15	1.39	-0.05	1.39	1
RLDA_14	1.33	0	1.33	1
CMMR_13	1.31	-0.06	1.31	1
STIU_15	1.3	-0.68	1.46	0.89
PLTN_13	1.23	-0.14	1.23	0.99
UBRT_14	1.22	-0.57	1.35	0.91
STIU_14	1.19	-0.23	1.21	0.98
STNP_14	1.18	-0.42	1.26	0.94
FLRT_14	1.16	-0.4	1.23	0.95
BVPR_14	1.12	-0.51	1.24	0.91
UBRT_15	1.11	-0.62	1.27	0.87
MAMB_12	1.1	-0.42	1.18	0.94
UBRT_13	1.09	-0.54	1.22	0.9
STNP_15	1.07	-0.64	1.25	0.86
PLTN_12	1.05	-0.54	1.18	0.89
RLDA_12	1.05	-0.18	1.06	0.99
CFLD_13	1.02	-0.21	1.04	0.98
MAMB_14	1.02	0.16	1.04	0.99
BVPR_12	1	-0.46	1.1	0.91
UBRT_16	1	-0.18	1.02	0.98
BVPR_16	0.96	-0.16	0.97	0.99
BVPR_13	0.95	-0.51	1.08	0.88
STNP_13	0.88	-0.12	0.89	0.99
BVPR_15	0.87	-0.36	0.94	0.92
STIU_16	0.85	0.5	0.99	0.86
LDRN_13	0.82	0.19	0.84	0.97
CMMR_12	0.8	-0.58	0.99	0.81
LDRN_16	0.8	-0.14	0.81	0.99
MRDN_14	0.75	0.92	1.19	0.63
LDRN_14	0.74	-0.21	0.77	0.96
CMMR_16	0.71	0.07	0.71	1
MRDN_12	0.68	0.06	0.68	1
CFLD_12	0.63	0.48	0.79	0.8

continue

continuation

STIU_13	0.58	0.52	0.78	0.75
TOLD_13	0.56	0.41	0.7	0.81
MRDN_16	0.48	-0.13	0.5	0.96
LDRN_15	0.44	0.1	0.45	0.98
FLRT_15	0.37	0.43	0.57	0.65
STNP_12	0.36	-0.25	0.44	0.82
PLTN_16	0.35	0.31	0.47	0.75
PLTN_14	0.33	0.7	0.77	0.43
TOLD_12	0.27	0.4	0.48	0.56
FLRT_12	0.22	0.27	0.35	0.62
PLTN_15	0.06	0.66	0.66	0.09
CFLD_15	0.05	0.62	0.62	0.07
CFLD_14	0.03	0.67	0.67	0.04
MAMB_13	-0.04	0.12	0.13	-0.31
CCVL_15	-0.38	0.64	0.74	-0.51
CCVL_13	-0.44	0.02	0.44	-1
CCVL_16	-0.68	0.06	0.68	-1
CFLD_16	-0.79	0.25	0.83	-0.95
TOLD_16	-0.82	0.5	0.96	-0.85
RLDA_13	-0.99	0.69	1.21	-0.82

*An environment is composed of a location and a year (for example RLDA_15 = Rolândia, in the year 2015). The locations are Rolândia (RLDA), Marechal Cândido Rondon (MRDN), Campo Mourão (CMMR), Santa Teresinha de Itaipu (STIU), Palotina (PLTN), Ubiratã (UBRT), Sertanópolis (STNP), Floresta (FLRT), Bela Vista do Paraíso (BVPR), Mambore (MAMB), Cafelândia (CFLD), Londrina (LDRN), Toledo (TOLD), and Cascavel (CCVL).

Table 5

Numerical values for the environment projections on the average environment axis (AEA; AEC_X) and their distances to the AEA (AEC_Y), and numerical values of the discriminating ability (vector length) and representativeness of the environments for edaphoclimatic region (ECR) 202

Environments*	AEC_X	AEC_Y	Vector Length	Correlation with AEA
FRCA_16	1.12	-0.33	1.17	0.96
NVRI_13	0.8	0.8	1.14	0.71
FRCA_13	0.68	0.65	0.94	0.72
NVRI_16	0.49	-1.01	1.12	0.43
IPOR_12	0.25	0.89	0.93	0.27
NVRI_12	0.15	0.62	0.64	0.24
NVRI_14	0.09	-1.19	1.19	0.08
IPOR_15	-0.01	-0.65	0.65	-0.02
FRCA_12	-0.14	-0.7	0.71	-0.19
NVRI_15	-0.21	0.92	0.94	-0.23

*An environment is composed of a location and a year (for example FRCA_16 = Francisco Alves, in the year 2016). The locations are Francisco Alves (FRCA), Naviraí (NVRI), and Iporã (IPOR).

Table 6
Numerical values for the environment projections on the average environment axis (AEA; AEC_X) and their distances to the AEA (AEC_Y), and numerical values of the discriminating ability (vector length) and representativeness of the environments for edaphoclimatic region (ECR) 204

Environments*	AEA_X	AEC_Y	Vector Length	Correlation with AEA
PTPR_16	1.48	-0.54	1.57	0.94
DOUR_13	1.41	-0.58	1.53	0.92
DOUR_15	1.25	-0.61	1.39	0.9
MRCJ_16	1.24	-0.33	1.29	0.97
PTPR_13	1.23	0.16	1.24	0.99
MRCJ_13	0.96	0.12	0.96	0.99
MRCJ_14	0.95	0.17	0.96	0.98
DOUR_14	0.92	0.2	0.94	0.98
DOUR_16	0.89	-0.36	0.96	0.93
MRCJ_15	0.65	-0.42	0.77	0.84
PTPR_15	0.58	0.07	0.58	0.99
SDLD_14	0.36	0.44	0.57	0.63
PTPR_14	0.14	0.54	0.56	0.25
PTPR_12	0.1	-0.19	0.22	0.45
SDLD_15	0.01	-0.01	0.02	0.65
MRCJ_12	-0.04	0.71	0.71	-0.05
DOUR_12	-0.28	0.64	0.7	-0.4

*An environment is composed of a location and a year (for example PTPR_16 = Ponta Porã, in the year 2016). The locations are Ponta Porã (PTPR), Dourados (DOUR), Maracaju (MRCJ), and Sidrolândia (SDLD).

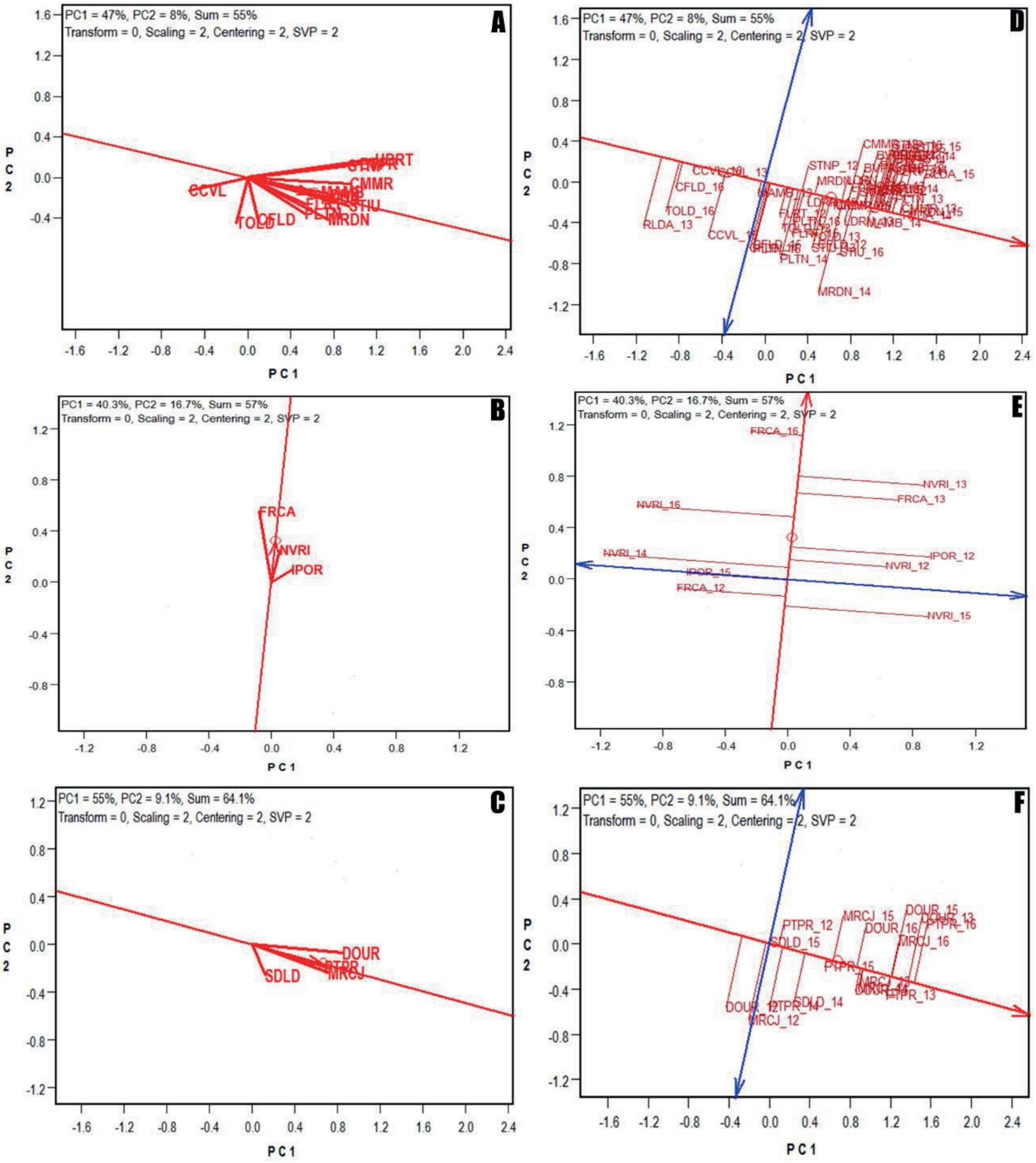

Figure 3. The GGL (genotype main effects plus genotype × location interaction) + GGE (genotype main effects plus genotype × environment interaction) biplot for macro-region 2 based on soybean crop data from 2012 to 2016 (201-A, 202-B, 204-C). The biplot form displays the environment's ability to select for G vs. GE (genotypic effect vs. genotype × environment effects) for macro-region 2 soybean crops (201-D, 202-E, 204-F). A Tester is composed of a location and a year (for example RLDA_13 = Rolândia, in the year 2013). The environments are: Rolândia (RLDA), Marechal Cândido Rondon (MRDN), Campo Mourão (CMMR), Santa Teresinha de Itaipu (STIU), Palotina (PLTN), Ubiratã (UBRT), Sertanópolis (STNP), Floresta (FLRT), Bela Vista do Paraíso (BVPR), Mambore (MAMB), Cafelândia (CFLD), Londrina (LDRN), Toledo (TOLD), Cascavel (CCVL), Francisco Alves (FRCA), Naviraí (NVRI), Iporã (IPOR), Ponta Porã (PTPR), Dourados (DOUR), Maracaju (MRCJ), and Sidrolândia (SDLD).

In ECR 202 (Figure 3B), location NVRI presented the highest r_g with the ECR, being the most representative environment. However, we observed greater consistency in the results for location FRCA, due to the longer vector. In the G vs. GE analysis, it was observed that IPOR_15, FRA_12, and NVRI_15 (Figure 3E) presented negative values for AEC_X (Table 5). In addition, the environment NVRI presented long vectors in all years, being inefficient in selecting genotypes by the G effect.

For ECR 204 (Figure 3C), locations PTPR and MRCJ presented the smallest angles with AEA and, therefore, the highest r_g, which makes them the most representative locations for this ECR. Location SDLD had the smallest vector and largest angle in relation to AEA (lower r_g), showing results that were both, inconsistent and unrepresentative. In the G vs. GE analysis (Figure 3F), environments DOUR_12 and MRCJ_12 were positioned on the left side of the AEC, with negative values of AEC_X (Table 6) and small $r_g h$ values, thus, showing a low capacity for selecting better genotypes. PTPR_16 was positioned on the right side of the biplot, indicating a high capacity for selection of superior genotypes. In addition, PTPR_13 also showed a high capacity to select superior genotypes, with greater ability to select genotypes by G effect, since it was characterized by a short vector, with an AEC_Y value of 0.16 (Table 6).

Discussion

The correct choice of test locations for conducting VCU trials is extremely important for breeding programs. Selected locations should maintain a similar pattern of genotypes responses over a timescale of years (consistency); they should be representative, and be able to discriminate genotypes (Yan et al., 2011; Luo et al., 2015; Yan, 2015). Thus, sites with high r_g values with the average environment and with long vectors (i.e., discriminative and consistent) should be retained in the set of locations used for VCU trials. In addition, locations at which genotypes can be selected by G effect are desirable, since there will be a lower effect from the environment on selection than at other locations. Thus, representative and discriminant locations, with a high capacity for selecting genotypes by G effect, can be considered ideal for conducting trials for the selection and recommendation of new cultivars (Yan et al., 2011; Yan, 2014). Moreover, when a wide MET is available, it is recommended that locations are included which encompass different levels of representativeness, i.e., that are representative of the TR, and at the same time reveal differentiated responses in the genotypes (Yan, Frégeau-Reid, Martin, Pageau, & Mitchell-Fetch, 2014). Furthermore, as well as identifying the best test locations, GGE biplot analyses also allow the identification of inappropriate (non-informative) locations (Krishnamurthy et al., 2017). Locations with similar patterns of representativeness and high r_g can generate redundant information (Yan et al., 2015); this behavior was observed in all ECRs. However, greater location similarity was observed in ECRs 102, 201, and 204 than in other locations.

For M1, locations TPJR and CMPV in ECR 102 (Figure 2B) may be substituted or even excluded from the MET. These locations presented poor results, were inefficient in selecting superior genotypes, and redundant in response to genotypes, even though they were representative of the TR. However, CNDR and SLGO were the least representative of the TR and can be excluded, because they do not contribute to the selection of superior genotypes in this region. Non-representative locations, such as CNDR and SLGO, or those with short and similar vectors to other locations (redundant and non-consistent/discriminant), such as TPJR and CMPV, should be removed from the MET, or replaced (Yan, 2015; Dia et al., 2016).

Regarding M2, locations in ECR 201 (Figure 3A) could be excluded, since they did not represent this ECR. These locations are inefficient for the selection of superior genotypes, they have low representativeness, select genotypes by GE effect (Figure 3D), and present similar results. Thus, three groups can be formed where sites can be excluded from ECR 201: Group 1 (CCVL, TOLD, and CFLD); Group 2 (UBRT, BVPR, and STNP); and Group 3 (MRDN, PLTN, FLRT, STIU, MAMB, CMMR, RLDA, and LDRN). Exclusion is justified by the high r_g between environments, which getting similar results between environments in these groups. In this sense, the exclusion of one or two locations per group would be justified, meaning these groups would continue to be represented, although by fewer locations. Similar environments result in redundant genotype responses, making the MET network inefficient (Yan, 2014; Yan et al., 2015).

In ECR 202, NVRI was the most representative between the locations evaluated (Figure 3B), presenting the highest genetic correlation with the ECR. This result indicates that the behavior of genotypes at this location can be extrapolated to the entire region. However, the long vectors in all years, indicated that this location is poor in the selection of genotypes by G effect (Figure 3E). In ECR 204 (Figure 3C), locations PTPR and MRCJ presented similar patterns of consistency and representativeness and were redundant (Yan et al., 2015). Thus, excluding one of these locations from the MET network can reduce breeding program costs. Further, SDLD can also be excluded, because it presented low consistency and moderate representativeness in this ECR.

It is recommended that the MET network includes locations that encompass different levels of representativeness, i.e., which are representative of the ME and, at the same time, reveal differentiated responses in the genotypes, to increase selection accuracy. Thus, non-representative or inconsistent locations must be replaced by others that are more informative, to maximize selection efficiency and reduce breeding program costs.

Conclusions

Representative and discriminating locations getting consistent results between years were identified, making it possible to select and recommend superior genotypes efficiently. Non-representative and redundant sites were also identified, which can be excluded from or replaced in the MET network, to increase selection efficiency and reduce costs. This was possible because the GGL + GGE analysis is superior to the year-by-year analyses and provides accurate results, even with missing data.

For M1, the locations most recommended for conducting VCU trials are Cachoeira do Sul (ECR 101); Ronda Alta, Passo Fundo, Santa Bárbara do Sul, and Ciríaco (ECR 102); and Castro (ECR 103). For M2, the locations recommended for conducting trials are Rolândia, Marechal Cândido Rondon, Campo Mourão, Santa Terezinha de Itaipu, Palotina, Floresta, and Londrina (ECR 201); Naviraí (ECR 202); and Ponta Porã and Maracajú (ECR 204).

References

Dia, M., Wehner. T. C., Hassell, R., Price, D. S., Boyhan, G. E., Olson, S.,… Juarez, B. (2016). Value of locations for representing mega-environments and for discriminating yield of watermelon in the US. *Crop Science, 56*(4), 1726-1735. doi:10.2135/cropsci2015.11.0698

Kaster, M., & Farias, J. R. B. (2012). *Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja: terceira aproximação.* Londrina: EMBRAPA Soja. Recuperado de http://ainfo.cnptia.embrapa.br/digital/bitstream/item/54939/1/Doc-330-OL1.pdf

Khatun, H., Islam, R., Anisuzzaman, M., Ahmed, H. U., & Haque, M. (2015). GGE biplot analysis of genotype × environment interaction in rice (*Oryza sativa* L.) genotypes in Bangladesh. *Scientia Agriculturae, 12*(1), 34-39. doi: 10.15192/PSCP.SA.2015.12.1.3439

Krishnamurthy, S. L., Sharma, P. C., Sharma, D. K., Ravikiran, K. T., Singh, Y. P., Mishra, V. K.,… Singh, R. K. (2017). Identification of mega-environments and rice genotypes for general and specific adaptation to saline and alkaline stresses in India. *Scientific Reports, 7*(7968), 1-14. doi:10.1038/s41598-017-08532-7

Luo, J., Pan, Y., Que, Y., Zhang, H., Grisham, M. P., & Xu, L. (2015). Biplot evaluation of test environments and identification of mega-environment for sugarcane cultivars in China. *Scientific Reports, 5*(15505), 1-11. doi: 10.1038/srep15505

Yan, W. (2001). GGEbiplot – A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal, 93*(5), 1111-1118. doi: 10.2134/agronj2001.9351111x

Yan, W. (2014). *Crop variety trials: data management and analysis.* Chichester, West Sussex, UK: Wiley Blackwell.

Yan, W. (2015). Mega-environment analysis and test location evaluation based on unbalanced multiyear data. *Crop Science, 55*(1), 113-122. doi:10.2135/cropsci2014.03.0203

Yan, W. (2016). Analysis and handling of G × E in a practical breeding program. *Crop Science, 56*(5), 2106-2118. doi:10.2135/cropsci2015.06.0336

Yan, W., Frégeau-Reid, J., Martin, R., Pageau, D., & Mitchell-Fetch, J. (2015). How many test locations and replications are needed in crop variety trials for a target region? *Euphytica, 202*(3), 361-372. doi: 10.1007/s10681-014-1253-7

Yan, W., & Holland, J. B. (2010). A heritability-adjusted GGE biplot for test environment evaluation. *Euphytica, 171*(3), 355-369. doi: 10.1007/s10681-009-0030-5

Yan, W., Pageau, D., Frégeau-Reid, J., & Durand, J. (2011). Assessing the representativeness and repeatability of test locations for genotype evaluation. *Crop Science, 51*(4), 1603-1610. doi:10.2135/cropsci2011.01.0016