Environmental Stratification in Trials of Unbalanced Multiyear Soybean (Glycine max (l.) Merril) via the Integration of GGE Biplot Graphs and Networks of Environmental Similarity

Fernanda Cupertino
UFV https://orcid.org/0000-0002-0748-8905
Francisco Charles Santos Silva (fcsantossilva-ma@hotmail.com)
Universidade Estadual do Maranhão
Pedro Crescêncio Souza Carneiro
Universidade Federal de Vicosa
Luiz Alexandre Peternelli
Universidade Federal de Vicosa
Leonardo Lopes Bhering
Universidade Federal de Vicosa
Felipe Lopes da Silva
Universidade Federal de Vicosa

Research Article

Keywords: multiyear trials, edaphoclimatic region, genotype–environment interaction, GGE biplot analysis, grain yield, multi-environment trials

Posted Date: November 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1027523/v1

Abstract

Genotype x enviroment (GE) interaction can difficult soybean breeding programs to atieve the aim of obtain more productive cultivars. Enviroment stratification is a way to circunvent this problem. This work aimed to gather GGE Biplot graphs of a network of trials unbalance multiyear soybean via matrices of coincidence and networks of enviroment to optimize environmental stratification. Data from an experimental network of 43 trials was used, these experiments were implanted during the crop seasons of 2011/12, 2012/13, 2013/14 and 2015/16 in Brazil. The GE interaction were statistically significant for all 43 trials. The step by step of our analses was: GGE Biplots graphs were obtained; the enviroment coincidence matrices were calculated; the values of matrices were used for to obtain the networks of environmental similarity. The study demonstrated that by the method was possible to identify, using unbalanced multiyear data, the formation of four mega-environments. Therefore, integrating GGE Biplot graphs and networks of environmental similarity is an efficient method to optimize a soybean program by environment stratification.

1 Introdution

The main aim of a soybean breeding program (*Glycine max* (L.) Merril) is the obtainment of more productive cultivars. However, this feature is controlled by many genes and it is highly influenced by the environment (E) and by the genotype x environment (GE) interaction. The GE interaction may cause divergences in the selection of genotypes and influence the gain of selection, which hinders the work of breeders, interfering in the selection and recommendation of superior cultivars.

There are two ways to circumvent the effects of the GE interaction: through the identification of more adapted and/or stable genotypes; and through environmental stratification, by forming mega-environments. Environmental stratification consists of subdividing environmentally heterogeneous areas into areas which are environmentally more homogeneous. Through stratification, it is possible to make decisions regarding the elimination and/or replacement of environmentally redundant planting sites with others with properties not yet sampled in the set of environments, optimizing the network of trials of the breeding program. This way, a better allocation of resources from the breeding program will be promoted, without loss of efficiency or precision in the selection process (Cruz & R 2007; Hühn & Truberg 2002; Oliveira; et al. 2007).

The GGE Biplot method (*Genotype Main Effect Plus Genotype x Environment Interaction*) (Yan et al. 2000), used for the identification of mega-environments, groups the additive effect of the genotype with the multiplicative effect of the GE interaction subjecting them to the analysis of principal components (PC) and provides graphic information on the GE interaction, which allows for the identification of different mega-environments. In the GGE methodology, the cosine of the angle formed between two environments in the graph corresponds to the genetic correlation between them. Other kinds of biplots do not have this property (Yan et al 2007), which makes this methodology more efficient if compared to other techniques based on biplots (Silva et al. 2008; Yan 2011). In addition, the GGE Biplot model explores the GE interaction more efficiently, allowing for greater accuracy in the identification of mega-environments. The greatest merit of the GGE Biplot methodology is when a large number of genotypes is tested in several environmental conditions and when the GE interaction standard is more complex (Silva & Benin 2012).

However, although the GGE method is widely used, it allows for, at most, 30% of the missing/unbalanced data (Woyann et al. 2017) and requires repeatability of the genotypes and of the sites throughout the crop seasons. In turn, multiyear trials, like the trials of value for cultivation and use (VCU), have a large amount of data and are usually unbalanced to the detriment of changes of the sites and/or genotypes tested throughout the years. Keeping this in mind, it becomes necessary to adapt the GGE Biplot method so that it is employed in multiyear data with more than 30% of unbalance.

Even with a high incidence of unbalance, an option to circumvent this problem is the use of matrices of environmental coincidence, which, on the whole, will indicate how many times a pair of environments was grouped in the same mega-environment in each crop season, which ensures more reliability in the stratification. For the matrices to be graphically represented, hence making it easier the identification of mega- environments, they can be represented by means of networks of environmental similarity. In this sense, this study aimed at gathering GGE Biplot graphs of a network of trials of unbalance multiyear soybean via matrices of coincidence and networks of environmental similarity to optimize environmental stratification.

2 Material And Methods

To carry out the methodology proposed, data of grain yield of soybean lines and of cultivars were used, which were kindly provided by GDM Seeds (GDM Genética do Brasil S.A.). These data are from an experimental network of 43 trials of VCU implanted during the crop seasons of 2011/12, 2012/13, 2013/14, and 2015/16 (supplementary material 1). The trials were implanted under an experimental block design with randomized treatments and with three replications. The parcels were formed by four lines of 5 meters. Spacing between the lines and culture handling followed the technical recommendations for soybean cultivation in each planting site.

Twenty-three municipalities of the states of Mato Grosso do Sul, Paraná, São Paulo and Santa Catarina (Table 1) formed the environments tested; however, it is worth pointing out that not all the environments were in every VCU trial (Table 2), and that for each trial a set of approximately 30 genotypes were evaluated, of which most of these genotypes were not repeated between the trials evaluated. Therefore, the data used had more than 30% of unbalance.

Table 1
– Plantations sites and their respective geographic coordinates.

State	Municipalities	Altitude	Latitude	Longitude
Mato Grosso do Sul	Dourados	430	22° 13' 18" S	54° 48' 23" O
	Maracaju	384	21º 36' 52" S	55º10'06" O
	Naviraí	362	23° 3' 55" S	54° 11' 26" O
	Ponta Porã	755	22° 32' 11" S	55° 43' 36" O
	Sidrolândia	483	20° 55' 54" S	54° 58' 10" O
Paraná	Bela Vista Do Paraíso	611	22° 59' 41" S	51° 11' 9" O
	Cafelândia	563	24° 37' 5" S	53° 19' 18" O
	Campo Mourão	596	24° 2' 46" S	52° 23' 2" O
	Cascavel	782	24° 57' 20" S	53° 27' 19" O
	Floresta	392	23° 35' 56" S	52° 04' 51" O
	Francisco Alves	323	24° 3' 50" S	53° 50' 21" O
	Iporá	584	16° 26' 29" S	51° 7' 11" O
	Londrina	550	23° 17' 34" S	23° 55' 46 O
	Mamborê	763	24° 16' 30" S	52° 30' 41" O
	Marechal Cândido Rondon	420	24° 33' 24" S	54° 3' 24" O
	Palotina	341	24° 16' 54" S	53° 50' 25" O
	Rolândia	736	23° 18' 38" S	51° 22' 10" O
	Santa Terezinha De Itaipu	284	25° 26' 56" S	54° 24' 26" O
	Toledo	550	24° 43' 12" S	53° 44' 36" O
	Ubiratã	507	24° 33' 18" S	52° 58' 40" O
Santa Catarina	Abelardo	771	26° 34' 2" S	52° 20' 2" O
São Paulo	Assis	546	22º 39' 42" S	50º 24' 44" O
	Florínea	360	22º 54' 12" S	50º 44' 16" O

Table 2
– Municipalities present in the trials of each crop season.

Municipalities per crop season			
2011/12	**2012/13**	**2013/14**	**2015/16**
Abelardo	Bela Vista do Paraíso	Assis	Assis
Bela Vista do Paraíso	Cafelândia	Bela Vista do Paraíso	Bela Vista do Paraíso
Cafelândia	Campo Mourão	Cafelândia	Cafelândia
Campo Mourão	Cascavel	Cascavel	Campo Mourão
Dourados	Dourados	Dourados	Cascavel
Floresta	Floresta	Floresta	Dourados
Florinea	Francisco Alves	Londrina	Floresta
Iporá	Londrina	Mamborê	Francisco Alves
Mamborê	Mamborê	Maracaju	Londrina
Maracaju	Maracaju	Marechal Cândido Rondom	Maracaju
Marechal Cândido Rondon	Naviraí	Naviraí	Naviraí
Naviraí	Palotina	Palotina	Palotina
Palotina	Ponta Porã	Ponta Porã	Ponta Porã
Ponta Porã	Rolandia	Rolandia	Rolandia
Rolandia	Santa Terezinha de Itaipu	Santa Terezinha de Itaipu	Santa Terezinha de Itaipu
Toledo	Toledo	Toledo	Toledo
	Ubiratã	Ubiratã	Ubiratã

Due to the low repeatability of the genotypes evaluated between the trials in the same crop season and between crop seasons, the strategy of analyzing the trials separately was adopted. This way, joint analyses of variance (ANOVA) were carried out for each trial, according to the statistical model described in the equation:

$Y_{ijk} = \mu + B/A_{jk} + G_i + A_j + GA_{ij} + e_{ijk}$

Where:

Y_{ijk} = observation of the i-th genotype evaluated in the k-th block of the j-th environment;

μ = general mean of the experiments;

B/A_{jk} = effect of block k within the j-th environment;

G_i = effect of the i-th genotype considered fixed;

A_j = effect of the j-th environment considered random;

GA_{ij} = random effect of the interaction between genotype i and environment j; and

e_{ijk} = random error associated with observation Y_{ijk}.

After carrying out the ANOVA for each trial in order to verify the GE interaction, analyses which aim to identify mega-environments were carried out. These analyses were divided into four steps (Figure1).

In the first step, GGE Biplot graphs were plotted for each trial and for each crop season. The following model was considered (Yan 2001):

$Y_{ij} - \bar{Y}_j = \xi_{i1}\square_{j1} + \xi_{i2}\square_{j2} + \square_{ij}$

Where:

Y_{ij} = average yield of genotype i on site j;

$\bar{Y}_j$ = average yield of all genotypes on site j;

ξ_{i1} e ξ_{i2} = singular values of PC1 and PC2, respectively, for genotype i;

$\square_{j1}$ = singular values of PC1 and PC2, respectively, for environment j;

$\square_{ij}$ = error associated to the bidimensional model, that is, the percentage of the G + GxE effects not explained by the first two PCs;

$\xi_{in} = \lambda_n\ \xi_{in}$; and

$\square_{jn} = \lambda_n\ \square_{jn}$.

The graphs were plotted by means of the Genes program (Cruz 2013) along with the agridat and GGE biplot Gui packages (R Core Team 2017). In order to do so, the following parameters were used: *Transform* = 0 (without transformation); *Scaling* = 1 [data scaled by the standard deviation of the mean of genotypes within the environments (SD)]; *Centering* = 2 (Main effect of genotype + genotype × environment interaction (G + GE)) and *singular values partition* (SVP) = 2 (Focus on the environment). This way, the data were weighted according to the standard deviation of the environments, where the size of the vectors tends to be the same, and the environments have the same weight in the analysis of genotypes, making it easier the identification of environmental associations (Yan 2015; Yan & Tinker 2006).

The GGE Biplot graph is built with basis on the axis of the first two PCs, in which they show percentage information of the variation explained by each principal component. The minimum percentage of explanation recommended for GGE Biplot analysis is 70% (Yan et al. 2000). When, in a trial, the sum of the two PCs represented less than 70% of the total variation, the strategy of subdividing the mega-environments found was adopted; this way, the environments in this trial were partitioned in two groups according to the average environment axis, and later, for each one of these groups new GGE graphs were plotted; the procedure was repeated until at least 70% of explanation of the total variation of the data was obtained in all the biplots.

To carry out the second step of the strategy of stratification, each GGE Biplot graph was analyzed separately. Due note was taken of which pairs of environments were grouped together in the same mega-environment, which allowed for the formation of a matrix of coincidence for each VCU of each crop season. This matrix was formed by numbers 0 and 1; zero was used when the pair of environments was not in the same mega- environment, and 1 was used when the pair was in the same mega-environment. These matrices indicate how many times each pair of environments was grouped in the same mega-environment.

Thus, in step three, it was possible to gather all the information contained in all the biplots, considering individual crop seasons, groupings of three crop seasons or the grouping of all crop seasons. For a better view of these matrices, in step four the matrix of coincidence was represented by means of a network of environmental similarity. The network environmental similarity was carried out by means of the Genes software (Cruz 2013).

The environmental stratify cations that were formed, by means of the integration of the GGE Biplot graphs in networks of environmental similarity were discussed, firstly considering the global coincidence matrix, followed by the matrices grouped in trios of crop seasons and finally by the individual matrices. This strategy favored to group, in an effective way, the highest number of coinciding environments in all crop seasons, when the global matrix was used.

3 Results And Discussion

The analyses of variance demonstrated that the effects of the genotypes and of the

GE interaction were significant ($p \leq 0{,}01$) for all 43 trials (supplementary material 3), which suggests the existence of different mega-environments within the study region, such a fact that allows the use of the GGE Biplot methodology.

The use of the first two principal components did not explain at least 70% of the total variation of 21 of the trials evaluated. Thus, for these trials, the environments were partitioned in two groups according to the average environment axis (AEA), as recommended by (Zdziarski et al. 2019), and, later on, a new GGE Biplot analysis was carried out for each one of the groups.

At the end of the GGE biplot analyses, 59 graphs of stratification were obtained (supplementary material 2), being 24 for the 2011/12 crop season; 15 for the 2012/13 crop season; six for the 2013/14 one and 14 for the 2015/16 crop season. By means of these graphs, matrices of coincidence were obtained and represented through networks of similarity.

To carry out the global analysis, nine coinciding municipalities among the four crop seasons were used. Based on the analyses (Figure 4), the standard of grouping of the municipalities throughout the four years revealed the existence of four mega-environments. Municipalities that belong to the same mega-environment promote similar effects on the tested genotypes, reducing the effect of the GE interaction. In this sense, they can be considered environmentally redundant to represent the identified mega-environment.

From the information generated by the network of similarity, it was possible to observe the independence of the towns of Bela Vista do Paraíso and Rolândia in relation to the other towns. Neither are not part of a mega-environment, which demonstrates that they have different environmental features from the others. An important feature to be mentioned is that the trials performed in the town of Bela Vista do Paraíso were irrigated, while the others are carried out in dry conditions. Thus, for the region being studied, these are considered essential environments for the implementation of soybean trials. This makes the representativeness of the study region more efficient, ensuring greater reliability in the selection and recommendation of soybean cultivars. It was also possible to notice that the towns of Maracaju, Palotina and Mamborê are present in more than one mega-environment, which demonstrates that they can represent more than one mega-environment.

The GGE Biplot methodology has been used to direct the planning of breeding programs in several regions and situations (Marcolin & Vezzetti 2017). In the present work, it was possible to identify municipalities capable of optimizing the implementation of a network of multi-environment trials for macro-region two of soybeans in Brazil, as established by (Kaster 2012).

It is important to reiterate that, for the region to be well represented, the testing sites must represent the environmental heterogeneity of the region; thus, environmentally redundant municipalities that do not contribute to the representativeness of the entire region must be replaced or ruled out. It is also important to point out that the towns which are environmentally distinct from the others must be maintained; this way there will be a better representativeness in the region through the complementarity of these sites.

Keeping this in mind, the optimization of the network of trials for the region being studied can be employed. It was found that the four mega-environments identified have intercessions; in these intercessions, towns are allocated and they can represent the respective mega-environments. Thus, it is recommended that these trials be performed in four of the nine coinciding environments evaluated, which are: Palotina, Maracaju, Bela Vista do Paraíso and Rolândia. The towns of Bela Vista do Paraíso and

Rolândia were included because they are considered essential environments in the representativeness of the soybean region being studied, as they are different from the others. The towns of Palotina and Maracaju were included in the environment to carry out the trials because both belong to more than one mega-environment, and so they can represent them. Palotina represents mega-environments ME1 and ME2 and Maracaju represents mega-environments ME3 and ME4.

In the recommendation of the municipalities to represent the region being studied, by means of an analysis of network of similarity considering all the crop seasons evaluated and the coinciding towns between them, it was possible to achieve a 55,6 % decrease in the number of municipalities included in the evaluations. The resources that before were directed to all these towns that provide redundant information of genotypes can be reallocated to the inclusion of other trials in municipalities that have not been adopted in the trial network yet, and/or can be directed to the inclusion of more genotypes to be evaluated in the representative towns of this network of trials.

For the analyses to be conclusive, the GGE Biplot method requires at least three years of trials so that the formation of mega-environments is analyzed (Yan & Frégeau-Reid 2018). Thus, in addition to the joint analysis of the four crop seasons, analyses of crop seasons combined in trios were also carried out. This way, it is possible to view some standards and/or groupings that did not occur when the matrix of global environmental coincidence was used. The analyses were performed with the coinciding towns among the trio of the evaluated crop seasons.

In the first analysis considering three crop seasons simultaneously (2011/12, 2012/13 and 2013/14), the formation of four mega-environments formed by seven towns was observed (Figure 5a). When comparing this analysis to the global analysis of the crop seasons, it is observed that Mamborê, Palotina and Cafelândia remained together in the same mega-environment. However, Palotina remains together with Cafelândia and also with Toledo, forming other two mega-environments, which is consistent with the choice of Palotina to represent these towns in the global analysis.

The towns of Maracaju and Palotina are still part of more than one mega-environment, which reinforces their similarity to the towns belonging to all the mega-environments that these are part of. The towns of Rolândia and Bela Vista do Paraíso, like in the global analysis, were not allocated in a mega-environment, which allows us to confirm that they environmentally differ from the others, and were very important in the formation of the testing sites.

1. **(b)**

1. **(d)**

DOR = Dourados; NV = Naviraí; MCJ = Maracaju; BV = Bela Vista do Paraíso; RL = Rolândia; CF = Cafelândia; PLT = Palotina; TL = Toledo; MCH = Marechal Cândido Rondon; FRT = Floresta e MAB = Mamborê.

For the trio of crop seasons 2011/12, 2012/13 and 2015/16, eight mega-environments were identified (Figure 5b). For this analysis, the towns of Maracaju, Palotina, Cafelândia and Rolândia were present in more than one mega-environment. No town was considered discrepant with the others; all of them were allocated in a mega-environment. Thus, the towns of Maracajú, Cafelândia and Palotina could represent the other sites. Such an analysis emphasizes that the towns of Palotina and Maracaju are environmentally similar to more than one town and can represent them in these trials for this region.

By analyzing crop seasons 2011/12, 2013/14 and 2015/16 together, 11 coinciding towns were evaluated, from which nine mega-environments were formed (Figure 5c). For this analysis, three towns that had not been studied yet were added, which are: Floresta and Marechal Cândido Rondon in the state of Paraná, and Naviraí in the state of Mato Grosso do Sul.

No municipalities were considered independent from the others, all of them belong to a mega-environment. It is interesting to notice that, although they are geographically closer and belong to the same state, the towns of Naviraí, Maracaju and Dourados are not allocated in the same mega-environment, which demonstrates that a shorter linear distance between two sites does not necessarily imply environmental homogeneity between them.

The towns of Maracaju, Cafelândia, Rolândia and Mamborê are once again allocated in more than one mega-environment. The towns of Marechal Cândido Rondon and Floresta, being studied for the first time in this analysis, are also in more than one mega-environment; besides that, they are allocated in the same mega-environment, which allows us to conclude that they are environmentally redundant and that only one of these sites will be selected for the implementation of the experiments. For this case, Floresta is the town selected, since Marechal Candido Rondon can also be represented by Maracaju and does not represent the town of Dourados.

In the evaluation, where the crop seasons of 2012/13, 2013/14 and 2015/16 were analyzed, nine coinciding towns were involved, of which seven formed five mega-environments (Figure 5d). As in the global analysis, the towns of Bela Vista do Paraíso and Rolândia proved to be discrepant from the others, not belonging to a mega-environment. Once again, the towns of Maracaju and Palotina are included in more than one mega-environment, and can represent other towns.

These networks of similarity traced with three crop seasons together confirmed some standards that could be observed in the global analysis, and also allowed for the study of two towns that were not present in such an analysis. Thus, the analyses in trio can generate important information, besides being conclusive.

Because not all the towns that participated in the trials for this region could be studied in the global analysis and in the analyses in trio, individual analyses for each crop season were also carried out; hence, all towns could be studied. Although the individual analysis is not conclusive, it can help in decision-making and it can also serve to characterize a specific crop season. Apart from that, these individual analyses can help the breeder to view grouping standards of existing environments and if they are repeated throughout the crop seasons.

In the 2011/12 crop season, the VCU trials were evaluated in 15 municipalities, which were grouped in eight mega-environments (Figure 6a).

(a) (b)

(c) (d)

It can be noticed that the towns of Palotina, Maracaju, Mamborê, Rolândia, Marechal Candido Rondon and Floresta presented ambiguity, belonging to more than one mega-environment. The towns of Bela Vista, Iporá and Toledo, in the state of Paraná, do not belong to a group, which suggests that they have environmental features that are specific and different from the others. Thus, these towns may have been important for the complementarity of the environments tested in the experimental network.

In this analysis, it is also possible to observe that Dourados, Naviraí and Maracaju, which are in the same state and are physically closer to each other, were not allocated in the same mega-environment. The town of Abelardo, in the state of Santa Catarina and the towns of Palotina, Campo Mourão and Mamborê belong to the same group. The town of Florínea, in the state of São Paulo and the towns of Marechal Candido Rondon and Floresta, in the state of Paraná, are also allocated in the same mega-environment. This reinforces the fact that the closest towns and/or which are in the same state will not necessarily be environmentally similar.

For the 2012/13 crop season, 17 towns were evaluated, of which 13 formed 9 mega-environments (Figure 6b). The town of Ponta Porã (PP), in Mato Grosso do Sul, presented a significant coincidence with the towns of Bela Vista do Paraíso (BV) and Santa Terezinha de Itaipu (STP) and Toledo (TL), demonstrating that the environmental influence of this town on the evaluated genotypes is similar to the other cited towns. The same happens to the towns of Cafelândia, Toledo, Marechal Candido Rondon and Naviraí, which also present a significant coincidence with other towns belonging to other mega-environments.

The other towns, Sidrolândia (SI), Ubiratã (UB), Mamborê (MAB) and Francisco Alves (FA), in the state of Paraná, did not combine in one mega-environment, which indicates a possible divergence of environmental influence on the genotypes between them and the other towns belonging to the mega-environments formed for this crop season.

In the 2013/14 crop season, VCU trials were implanted in 18 towns, of which seven were grouped into three mega-environments (Figure 6c). It is noticed that, like in the previous crop season (2012/13), the towns of Dourados, Maracaju and Naviraí remain allocated in the same mega-environment. When evaluating the grouping obtained for the 2015/16 crop season, it was verified that of the 16 towns evaluated, eight formed six mega-environments (Figure 6d).

By means of the analyses carried out, it was possible to identify some stratification standards throughout the crop seasons. The towns of Naviraí (NV) and Maracaju (MCJ) remained in the same mega-environment in all the evaluated crop seasons. As for the towns of Bela Vista do Paraíso (BV), Toledo (TL) and Rolândia (RL), they appear far from the others, not remaining in any group in the crop seasons of 2011/12, 2013/14 and 2015/16.

4 Conclusions

Through the joint analysis of the nine coinciding environments between the crop seasons, it was possible to group GGE Biplot graphs of unbalanced multiyear data via matrices of coincidence and networks of environmental similarity, which made it possible to identify the formation of four mega-environments. The region under study can be represented by the towns of Palotina, Maracaju, Bela Vista do Paraíso and Rolândia, which reduces the number of towns to be evaluated and optimizes the trials in this region.

The towns of Rolândia and Bela Vista do Paraíso have environmental features that are different from the others. Therefore, they do not provide redundant information, and are important to form the environments tested for the selection and recommendation of the cultivars.

Declarations

ACKNOWLEDGMENTS

The authors thank GDM Seeds (GDM Genética do Brasil S.A.) for providing the multiyear data. The authors also thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial suport.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

References

Cruz, C. D. (2013). GENES - Software para análise de dados em estatística experimental e em genética quantitativa. *Acta Scientiarum - Agronomy*, *35*(3), 271–276. https://doi.org/10.4025/actasciagron.v35i3.21251

Cruz, C. D., & R, A. J. (2007). *Modelos Biométricos Aplicados ao Melhoramento Genético* (4th ed.). Editora UFV.

Hühn, M., & Truberg, B. (2002). Contributions to the analysis of genotype x environment interactions: Theoretical results of the application and comparison of clustering techniques for the stratification of field test sites. *Journal of Agronomy and Crop Science*, *188*(2), 65–72. https://doi.org/10.1046/j.1439-037X.2002.00549.x

Kaster, M. (2012). *Regionalização dos testes de Valor de Cultivo e Uso e da indicação de cultivares de soja - terceira Aproximação.* 231–235.

Marcolin, F., & Vezzetti, E. (2017). Novel descriptors for geometrical 3D face analysis. *Multimedia Tools and Applications*, *76*(12), 13805–13834. https://doi.org/10.1007/s11042-016-3741-3

Oliveira;, A. M. da S. de, Hamawaki, O. T., Neto, J. O. de O., Penna, J. C. V., Juliatti, F. C., Souza, S. A. de, Hühn, M., Truberg, B., Cruz, C. D., R, A. J., & Egazzi. (2007). Modelos Biométrocos Aplicados ao Melhoramento Genético. In *Journal of Agronomy and Crop Science* (4th ed., Vol. 188, Issue 2). Editora UFV. https://doi.org/10.1046/j.1439-037X.2002.00549.x

R, (CoreTeam). (2017). *R: A Language and Environment for Statistical Computing*. *2*, 1–12. https://www.r-project.org/

Silva, R. R., & Benin, G. (2012). Análises Biplot: conceitos, interpretações e aplicações. *Ciência Rural*, *42*(8), 1404–1412. https://doi.org/10.1590/s0103-84782012000800012

Silva, R. R., Benin, G., Silva, G. O. da, Marchioro, V. S., Almeida, J. L. de, & Gilvani Matei. (2008). Adaptabilidade e estabilidade de cultivares de soja em quatro épocas de semeadura no sul do Estado do Tocantins. *Revista Ceres*, *55*(1), 34–40.

Woyann, L. G., Benin, G., Storck, L., Trevizan, D. M., Meneguzzi, C., Marchioro, V. S., Tonatto, M., & Madureira, A. (2017). Estimation of missing values affects important aspects of GGE biplot analysis. *Crop Science*, *57*(1), 40–52. https://doi.org/10.2135/cropsci2016.02.0100

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. https://doi.org/10.2134/agronj2001.9351111x

Yan, Weikai. (2015). Mega-environment Analysis and Test Location Evaluation Based on Unbalanced Multiyear Data. *Crop Science*, *55*(1), 113–122. https://doi.org/10.2135/cropsci2014.03.0203

Yan, WeiKai. (2011). GGE biplot vs. AMMI graphs for genotype-by-environment data analysis. *Journal of the Indian Society of Agricultural Statistics*, *95*, 181–193. https://www.cabdirect.org/cabdirect/abstract/20113313412

Yan, Weikai, & Frégeau-Reid, J. (2018). Genotype by Yield∗Trait (GYT) Biplot: A Novel Approach for Genotype Selection based on Multiple Traits. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-26688-8

Yan, Weikai, Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, *40*(3), 597–605. https://doi.org/10.2135/cropsci2000.403597x

Yan, Weikai, Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science*, *47*(2), 643–655. https://doi.org/10.2135/cropsci2006.06.0374

Yan, Weikai, & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, *86*(3), 623–645. https://doi.org/10.4141/P05-169

Zdziarski, A. D., Woyann, L. G., Milioli, A. S., Zanella, R., Dallacorte, L. V, & Panho, M. C. (2019). *Mega - environment identification for soybean (Glycine max) breeding and production in Brazilian Midwest region*. *July 2018*, 336–347. https://doi.org/10.1111/pbr.12693

Figures

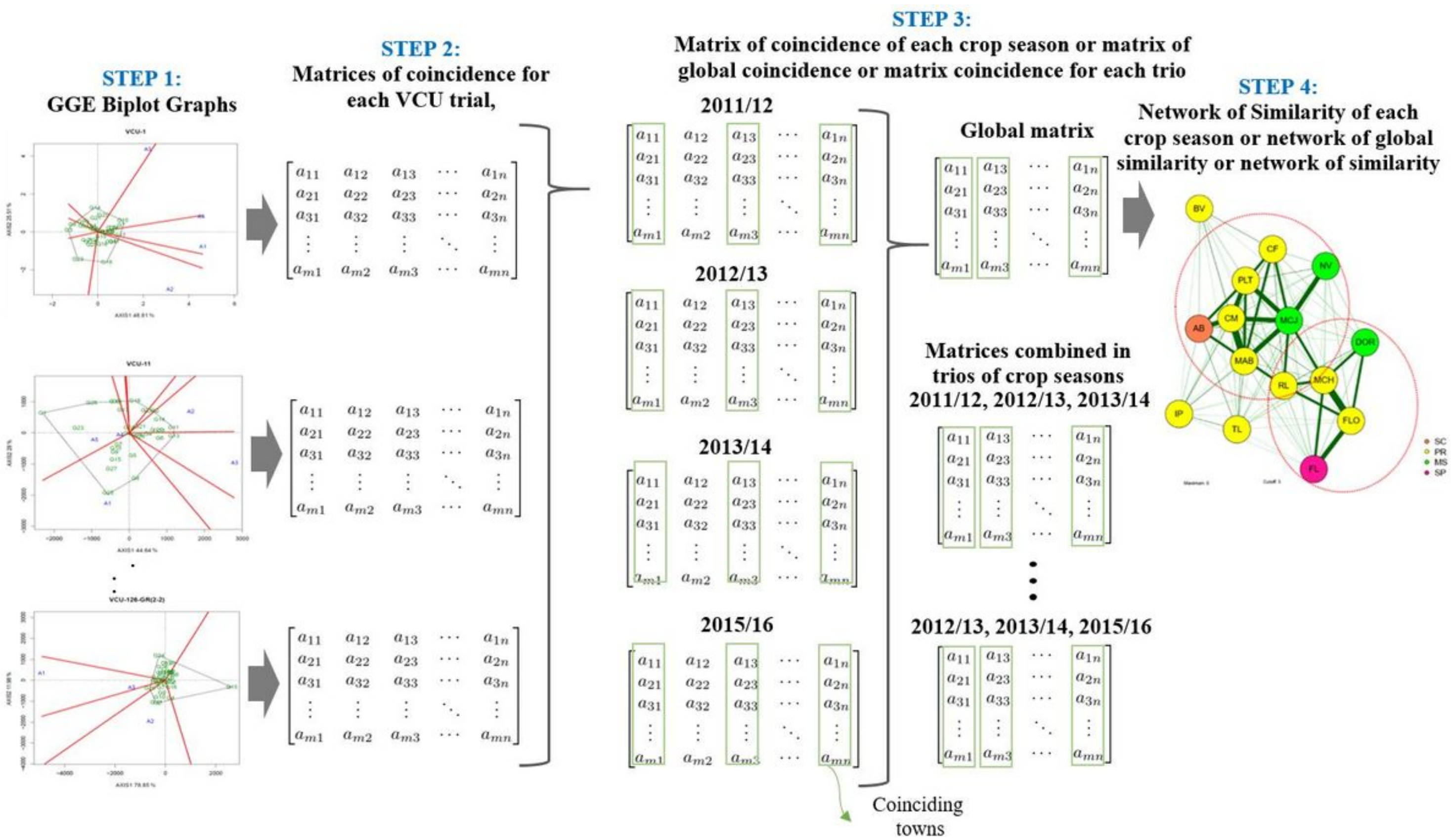

Figure 1

Flow chart of the stages of the statistical analyses for the formation of global and individual mega-environment of the crop seasons. The yellow circles are indicative of locations belonging to the state of Paraná; orange circles are indicative of locations belonging to the state of Mato Grosso do Sul; green circles are indicative of locations belonging to the state of Santa Catarina and pink circles are indicative of locations belonging to the state of São Paulo.

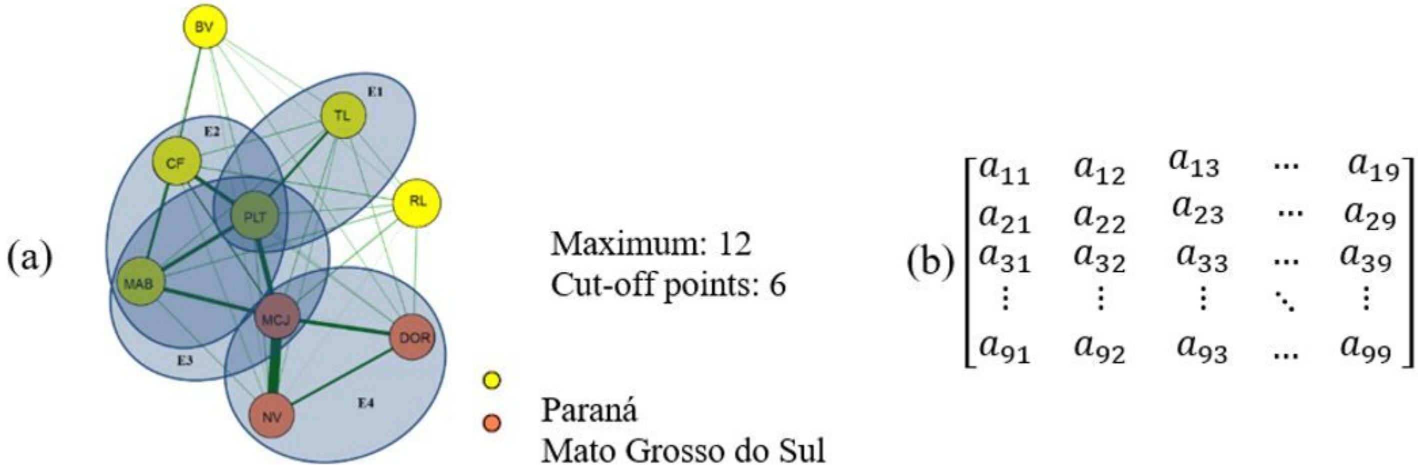

Figure 2

Network of global environmental similarity (a) and global matrix (b) for crop seasons 2011/12, 2012/13, 2013/14 and 2015/16. The yellow circles are indicative of locations belonging to the state of Paraná; orange circles are indicative of locations belonging to the state of Mato Grosso do Sul and blue circles indicate mega-environments. BV = Bela Vista do Paraíso; TL = Toledo; CF = Cafelândia; PLT = Palotina; RL = Rolândia; MAB = Mamborê; MCJ = Maracaju; NV = Naviraí; DOR = Dourados

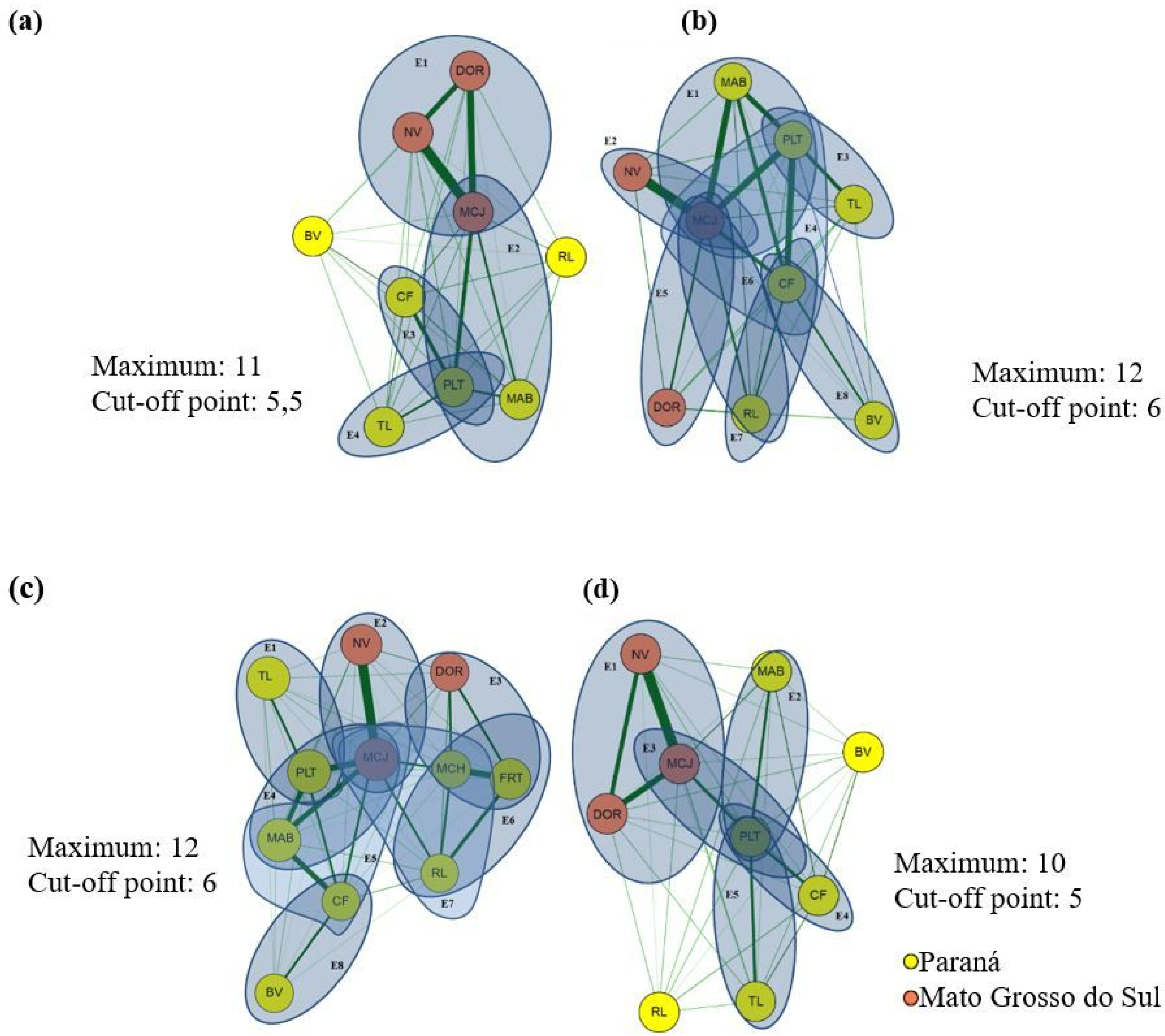

Figure 3

Network of joint environmental similarity for crop seasons (a) 2011/12, 2012/13 and 2013/14; (b) 2011/12, 2012/13 and 2015/16; (c) 2011/12, 2013/14 and 2015/16 and (d) 20012/13, 2013/14 and 2015/16. The yellow circles are indicative of locations belonging to the state of Paraná; orange circles are indicative of locations belonging to the state of Mato Grosso do Sul; blue circles indicate mega-environments. DOR = Dourados; NV = Naviraí; MCJ = Maracaju; BV = Bela Vista do Paraíso; RL = Rolândia; CF = Cafelândia; PLT = Palotina; TL = Toledo; MCH = Marechal Cândido Rondon; FRT = Floresta e MAB = Mamborê.

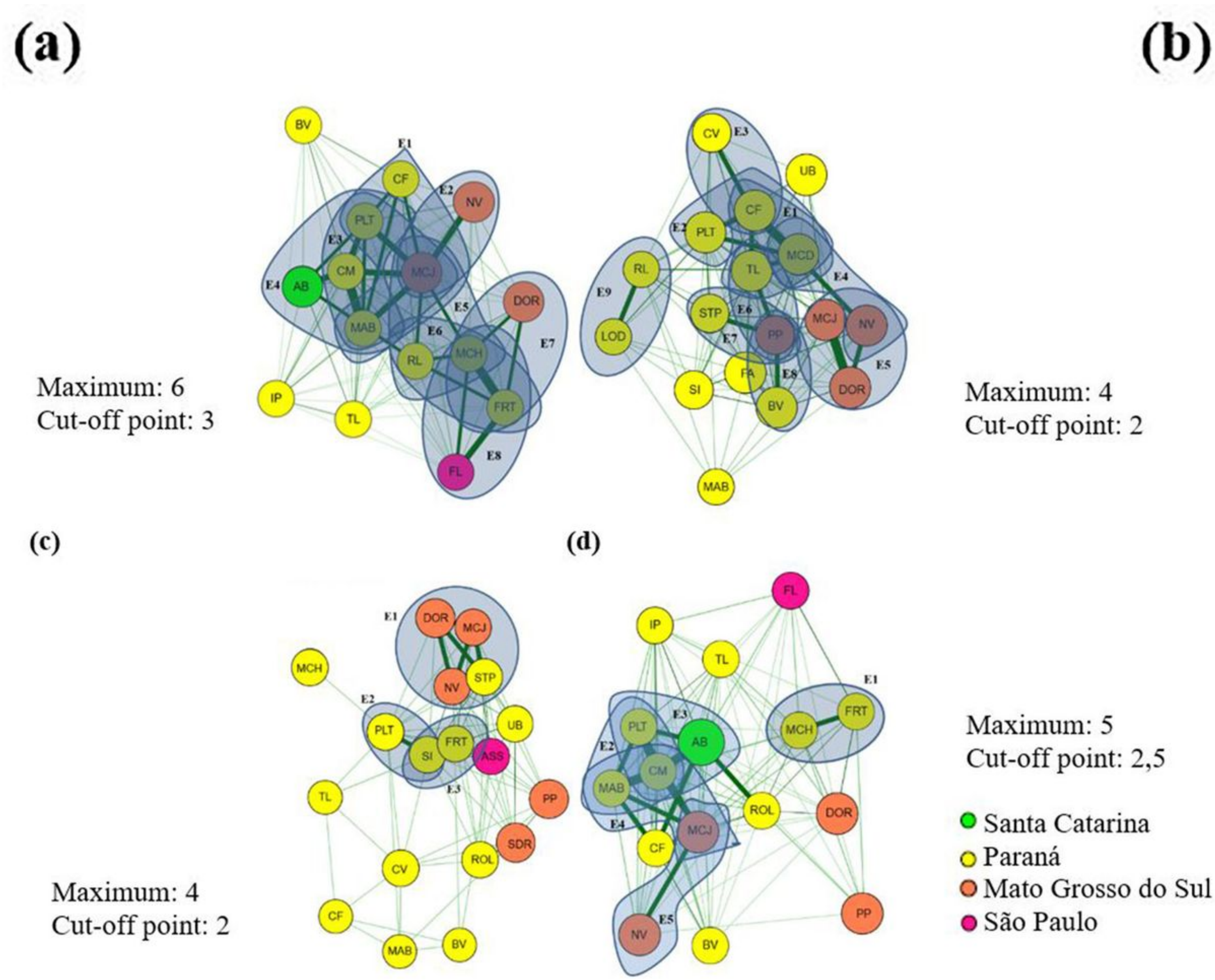

Figure 4

Network of environmental similarity for crop seasons (a) 2011/12; (b) 2012/13; (c) 2013/14 and (d) 2015/16. The yellow circles are indicative of locations belonging to the state of Paraná; orange circles are indicative of locations belonging to the state of Mato Grosso do Sul; gree circles are indicative of locations belonging to the state of Santa Catarina; pink circles are indicative of locations belonging to the state of São Paulo; blue circles indicate mega-environments. blues circles indicate mega-environments. DOR = Dourados; NV = Naviraí; MCJ = Maracaju; BV = Bela Vista do Paraíso; RL = Rolândia; CF = Cafelândia; PLT = Palotina; TL = Toledo; MCH = Marechal Cândido Rondon; FRT = Floresta; MAB = Mamborê; AB = Abelardo; CM = Campo Mourão; FL = Florínea e IP = Iporá

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYMATERIAL.docx