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 Rodrigues · Francisco Charles Santos Silva · Pedro Crescêncio Souza Carneiro · Luiz Alexandre Peternelli · Leonardo Lopes Bhering · Felipe Lopes da Silva

Received: 1 November 2021 / Accepted: 23 February 2022 / Published online: 13 May 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract Genotype × environment (GE) interaction can difficult soybean breeding programs to achieve the aim of more productive cultivars. Environment stratification is a way to circumvent this problem. However, multiyear data studies are difficult to work with, because they are, usually, unbalanced. GGE Biplot is an efficient method to find mega-environments, however, it allows for, at most, 30% of the unbalanced data. Thinking about how we could resolve this problem we came up with the idea of test a method that englobes GGE Biplot graphs, environment coincidence matrices and networks of environment. Wherefore, this work aimed to gather GGE Biplot graphs of a network of trials unbalance multiyear soybean via matrices of coincidence and networks of environment to optimize environmental stratification. Data from an experimental network of 43 trials was used, these experiments were implanted in 23 municipalities during the crop seasons of 2011/2012, 2012/2013, 2013/2014 and 2015/2016 in Brazil. The trials were implanted under an experimental block design with randomized treatments and approximately 30 genotypes were evaluated, of which most of these genotypes were not repeated between the trials evaluated. The GE interaction were statistically significant for all 43 trials. The step by step of our analyses was: GGE Biplots graphs were obtained; the environment coincidence matrices were calculated; the values of matrices were used to obtain the networks of environmental similarity. The study demonstrated that by the method was possible to identify, using unbalanced multiyear data, four mega-environments. The region under study can be represented by the municipalities of Palotina, Maracaju, Bela Vista do Paraíso and Rolândia. Therefore, integrating GGE Biplot graphs and networks of environmental similarity is an efficient method to optimize a soybean program by environment stratification.

Keywords Multiyear trials · Edaphoclimatic region · Genotype–environment interaction · GGE Biplot analysis · Grain yield · Multi-environment trials
Introduction

The main aim of a soybean breeding program (*Glycine max* (L.) Merril) is to obtain more yielding 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 (Cruz et al. 2007).

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 (Cruz et al. 2007). 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 (Hühn and 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 (Pelúzio 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 and 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), used in this present study, 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.

Analyzing how this problem could be solved, came up the idea of using 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 with networks of environmental similarity, which can represent the matrices graphically, hence making it easier the identification of mega-environments (Silva et al. 2020; Ramburan et al., 2012). 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.

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 2011/2012, 2012/2013, 2013/2014, and 2015/2016 (supplementary material 1). The trials were implanted under an experimental block design with randomized treatments and with three replications. The plots were formed by four lines of 5 m. 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
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      | 21° 36' 52" S    | 55° 10'06" 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            | Abielardo            | 771      | 26° 34' 2" S     | 52° 20' 2" O     |
| São Paulo                | Assis                | 546      | 22° 39' 42" S    | 50° 24' 44" O    |
|                           | Florínea             | 360      | 22° 54' 12" S    | 50° 44' 16" O    |

Table 2  Municipalities present in the trials of each crop season

Municipalities per crop season

| 2011/2012 | 2012/2013 | 2013/2014 | 2015/2016 |
|-----------|-----------|-----------|-----------|
| 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 | Cafelândia | Campo Mourão |
| Dourados  | Dourados  | Dourados  | Dourados  |
| Floresta  | Floresta  | Floresta  | Dourados  |
| Florínea  | Francisco Alves | Londrina | Floresta |
| Iporá     | Londrina  | Mamborê   | Francisco Alves |
| Mamborê   | Mamborê   | Maracaju  | Londrina  |
| Maracaju  | Maracaju  | Marechal Cândido Rondon | Maracaju |
| Marechal Cândido Rondon | Naviraí | Naviraí | Naviraí |
| Naviraí   | Palotina  | Palotina  | Palotina  |
| Palotina  | Ponta Porã| Ponta Porã| Ponta Porã |
| Ponta Porã| Rolândia  | Rolândia  | Rolândia  |
| Rolândia  | Santa Terezinha de Itaipu | Santa Terezinha de Itaipu | Santa Terezinha de Itaipu |
| Toledo    | Toledo    | Toledo    | Toledo    |
| Ubiratã   | Ubiratã   | Ubiratã   | Ubiratã   |
repeated between the trials evaluated. Therefore, the data used had more than 30% of unbalance.

2.1 Joint analyses of variance (ANOVA)

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 ith genotype evaluated in the kth block of the jth environment; \( \mu \) = general mean of the experiments; \( B/A_{jk} \) = effect of block k within the jth environment; \( G_i \) = effect of the ith genotype considered fixed; \( A_j \) = effect of the jth environment considered random; \( GA_{ij} \) = random effect of the interaction between genotype i and environment j; and \( e_{ijk} \) = random error associated with observation \( Y_{ijk} \).

Mega-environments

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 (Fig. 1).

Step 1 GGE Biplot Graphs.

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

\[ Y_{ij} - \bar{Y}_j = \xi_i \eta_j + \xi_i \eta_j + \epsilon_{ij} \]

where \( Y_{ij} \) = average yield of genotype i on site j; \( \bar{Y}_j \) = average yield of all genotypes on site j; \( \xi_i \) and \( \xi_i \) = singular values of PC1 and PC2, respectively, for genotype i; \( \eta_j \) and \( \eta_j \) = singular values of PC1 and PC2, respectively, for environment j; \( \epsilon_{ij} \) = error associated to the bidimensional model, that is, the percentage of the GxGxE effects not explained by the first two PCs;

\[ \epsilon_{in} = \lambda_n^{1/2} \epsilon_{in} \] and
The graphs were plotted by means of the Genes program (Cruz 2013) along with the agridat and GGEBiplotGUI packages (R CoreTeam 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 \(\times\) 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 and 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.

Step 2  Matrix of coincidence for each VCU trial.

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.

Step 3  Matrix of coincidence of each crop season or matrix of global coincidence or matrix coincidence for each trio.

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).

Step 4  Network of Similarity of each crop season or network of global similarity or network of similarity.

The environmental stratify 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 (Fig. 2).

Results and discussion

Global analysis

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 Dalló 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 analysis, 59 graphs of stratification were obtained (supplementary material 2), being 24 for the 2011/2012 crop season; 15 for the 2012/2013 crop season; six for the 2013/2014 one and 14 for the 2015/2016 crop season. By means of these graphs, matrices of coincidence were obtained and represented through networks of similarity (Fig. 3).

To carry out the global analysis, nine coinciding municipalities among the four crop seasons were used. Based on the analyses (Fig. 4), the standard of grouping of the municipalities throughout the

**Fig. 2** 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.

**Fig. 3** Network of global environmental similarity (a) and global matrix (b) for crop seasons 2011/2012, 2012/2013, 2013/2014 and 2015/2016. 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 = Mamboré; MCJ = Maracaju; NV = Naviraí; DOR = Dourados. (Color figure online)
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 (Dalló et al. 2019; Choudhary et al. 2019). 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 municipalities of Bela Vista do Paraíso and Rolândia in relation to the other municipalities. Neither are 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 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.

The GGE Biplot methodology has been used to direct the planning of breeding programs in several regions and situations (Marcolin and Vezzetti 2017). In this 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 municipalities 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 (Sousa et al. 2018; da Silva et al. 2021; Dalló et al. 2019).

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, municipalities 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 municipalities 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. Palotina and Maracaju were included in the environment to carry out the trials because both belong to more than one mega-environment, which demonstrates that they can represent more than one mega-environment. 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 municipalities 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 municipalities 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 municipalities of this network of trials. In this way, the breeding program optimizes resources and conducts its experiments more efficiently.

Trio analysis

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 and 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 municipalities among the trio of the evaluated crop seasons.

In the first analysis considering three crop seasons simultaneously (2011/2012, 2012/2013 and 2013/2014), the formation of four mega-environments by seven municipalities was observed (Fig. 4a). 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 municipalities in the global analysis. Maracaju and Palotina are still part of more than one mega-environment, which reinforces their similarity to the municipalities belonging to all the mega-environments that these are part 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.

For the trio of crop seasons 2011/2012, 2012/2013 and 2015/2016, eight mega-environments were identified (Fig. 4b). For this analysis, the municipalities of Maracaju, Palotina, Cafelândia and Rolândia were present in more than one mega-environment. No municipalities was considered discrepant with the others; all of them were allocated in a mega-environment. Thus, Maracajú, Cafelândia and Palotina could represent the other sites. Such an analysis emphasizes that 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 2011/2012, 2013/2014 and 2015/2016 together, 11 coinciding municipalities were evaluated, from which nine mega-environments were formed (Fig. 4c). For this analysis, three municipalities 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. The inclusion of this municipalities in the analysis were important to observe their behavior, as this was not possible in the global analysis. 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 municipalities 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. Maracaju, Cafelândia,
Rolândia and Mamborê are once again allocated in more than one mega-environment. 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 2012/2013, 2013/2014 and 2015/2016 were analyzed, nine coinciding municipalities were involved, of which seven formed five mega-environments (Fig. 4d). As in the global analysis, Bela Vista do Paraíso and Rolândia proved to be discrepant from the others, not belonging to a mega-environment. Once again, Maracaju and Palotina are included in more than one mega-environment and can represent other municipalities.

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 municipalities that were not present in such an analysis. Thus, the analyses in trio can generate important information, besides being conclusive.

Individual analysis

Because not all the municipalities 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 municipalities could be studied. Although the individual analysis is not conclusive, it can help in

---

**Fig. 5** Network of environmental similarity for crop seasons a 2011/2012; b 2012/2013; c 2013/2014 and d 2015/2016. 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; pink circles are indicative of locations belonging to the state of São Paulo; blue circles indicate mega-environments. Blues circles indicate mega-environments. Blues circles indicate mega-environments. DOR = Dourados; NV = Naviraí; MCI = 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. (Color figure online)
decision-making and it can also be used 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 2011/2012 crop season, the VCU trials were evaluated in 15 municipalities, which were grouped in eight mega-environments (Fig. 5a). It can be noticed that Palotina, Maracaju, Mamborê, Rolândia, Marechal Candido Rondon and Floresta presented ambiguity, belonging to more than one mega-environment. Bela Vista, Ipóra 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 municipalities 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. Abelardo, in the state of Santa Catarina and Palotina, Campo Mourão and Mamborê belong to the same group. Floreína, in the state of São Paulo and 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 municipalities and/or which are in the same state will not necessarily be environmentally similar.

For the 2012/2013 crop season, 17 municipalities were evaluated, of which 13 formed 9 mega-environments (Fig. 5b). Ponta Porã (PP), in Mato Grosso do Sul, presented a significant coincidence with Bela Vista do Paraíso (BV), 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 municipalities. The same happens to Cafelândia, Toledo, Marechal Candido Rondon and Naviraí, which also present a significant coincidence with other municipalities belonging to other mega-environments. The others municipalities, 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 municipalities belonging to the mega-environments formed for this crop season.

In the 2013/2014 crop season, VCU trials were implanted in 18 municipalities, of which seven were grouped into three mega-environments (Fig. 5c). It is noticed that, like in the previous crop season (2012/2013), Dourados, Maracaju and Naviraí remain allocated in the same mega-environment. When evaluating the grouping obtained for the 2015/2016 crop season, it was verified that of the 16 municipalities evaluated, eight formed six mega-environments (Fig. 5d).

By means of the analysis carried out, it was possible to identify some stratification standards throughout the crop seasons. Naviraí (NV) and Maracaju (MCJ) remained in the same mega-environment in all the evaluated crop seasons. As for 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 2011/2012, 2013/2014 and 2015/2016.

All these analyzes proved to be very important to contribute to the advancement of the knowledge of GE interaction analysis in unbalanced data. These analyzes allow us to visualize with greater background and precision field experiments and crops, making it possible to observe and solve problems that the GE interaction can bring to breeders, breeding companies and farmers.

Conclusions

The region under study can be represented by the municipalities of Palotina, Maracaju, Bela Vista do Paraíso and Rolândia. The municipalities 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. Our study demonstrated that integrating GGE Biplot graphs, matrices of environment and networks of environmental similarity is an efficient method to optimize a soybean program by environment stratification analyzing unbalanced data. We revealed that the number of municipalities used to implement the experiments can be reduced, since some of these cities are in the same mega-environment, which optimizes breeding programs by reducing costs and
improving the methodologies to obtain more yielding soybean cultivars.

Acknowledgements 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.

Funding This work was supported the Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES and the Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Choudhary M, Kumar B, Kumar P, Guleria SK, Singh NK, Khulbe R, Kamboj MC, Vyas M, Srivastava RK, Puttaramanaiik SD, Mahajan V, Rakshit S (2019) GGE biplot analysis of genotype × environment interaction and identification of mega-environment for baby corn hybrids evaluation in India. Indian J Genet 79(4):658–669. https://doi.org/10.31742/IJGPB.79.4.3

Cruz CD (2013) GENES–GENES—a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum—Agronomy 35(3):271–276. https://doi.org/10.4025/actasiagron.v35i3.21251

Cruz CD, Regazzi AJ (2007) Modelos Biométricos Aplicados ao Melhoramento Genético, 4th ed. Editora UFV

Dalló SC, Zdziarski AD, Woyann LG, Milioli AS, Zanella R, Conte J, Benin G (2019) Across year and year-by-year GGE biplot analysis to evaluate soybean performance and stability in multi-environment trials. Euphytica 215(6):1–12. https://doi.org/10.1007/s10681-019-2438-x

da Silva KJ, Teodoro PE, da Silva MJ, Teodoro LPR, Cardoso MJ, Godinho VDPC, Mota JH, Simon GA, Tardin FD, da Silva AR, Guedes FL, de Menezes CB (2021) Identification of mega-environments for grain sorghum in Brazil using GGE biplot methodology. Agron J 113(4):3019–3030. https://doi.org/10.1002/agj2.20707

Hühn M, Truberg B (2002) Contributions to the analysis of genotype × environment interactions: theoretical results of the application and comparison of clustering techniques for the stratification of field test sites. J Agron Crop Sci 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 pp 231–235

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

Oliveira, AMdaS de AMdaS, Hamawaki OT, Neto JOdeO, Penna JCV, Juliatti FC, de Souza SA, Hühn M, Truberg B, Cruz CD, R AJ, Egazzi (2007) Modelos Biométricos Aplicados ao Melhoramento Genético. In J Agron Crop Sci 4th ed., vol 188, Issue 2. Editora UFV. https://doi.org/10.1046/j.1439-037X.2002.00549.x

Pelúzio JM, Fidelis RR, Giongo P, Silva JCD, Cappellari D, Barros HB (2008) Adaptabilidade e estabilidade de cultivares de soja em quatro épocas de semeadura no sul do Estado do Tocantins

Ramburan S, Zhou M, Labuschagne MT (2012) Investigating test site similarity, trait relations and causes of genotype×environment interactions of sugarcane in the Midlands region of South Africa. Field Crop Res 129:71–80. https://doi.org/10.1016/j.fcr.2012.01.017

R (CoreTeam) (2017) R: a language and environment for statistical computing 2, 1–12. https://www.r-project.org/

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

Silva GN, Júnior ACS, Sant’Anna IC, Cruz CD, Nascimento M, Soares PC (2020) Similarity networks for the classification of rice as to adaptability and stability. Scielo. https://doi.org/10.1590/s0103-847820200000012

Sousa MBE, Damasceno-Silva KJ, Rocha MDM, Menezes Júnior JÂND, Lima LRL (2018) Genotype by environment interaction in cowpea lines using GGE biplot method. Revista Caatinga 31:64–71. https://doi.org/10.1590/1983-21252018v31n108rc

Woyann LG, Benin G, Storck L, Trevizan DM, Meneguzzi C, Marchioro VS, Tomatto M, Madureira A (2017) Estimation of missing values affects important aspects of GGE biplot analysis. Crop Sci 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. Agron J 93(5):1111–1118. https://doi.org/10.2134/agronj2001.9351111x

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

Yan W (2011) GGE biplot versus AMMI graphs for genotype-by-environment data analysis. J Indian Soc Agric Stat 95:181–193. https://www.cabdirect.org/cabdirect/abstract/2011331412

Yan W, Frégeau-Reid J (2018) Genotype by Yield*Trait (GYT) Biplot: a novel approach for genotype selection based on multiple traits. Sci Rep 8(1):1–10. https://doi.org/10.1038/s41598-018-26688-8

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

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot versus AMMI analysis of genotype-by-environment data. Crop Sci 47(2):643–655. https://doi.org/10.2135/cropsci2006.06.0374

Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: principles and applications. Can J Plant Sci 86(3):623–645. https://doi.org/10.4141/P05-169

Springer
Publisher’s Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.