

BRUNO GRESPAN LEICHTWEIS

**MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN PARA AGRUPAMENTOS
AMBIENTAIS E ESTUDOS DE INTERAÇÃO GENÓTIPOS POR AMBIENTES VIA
NORMAS DE REAÇÃO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

Orientador: Felipe Lopes da Silva

Coorientadores: Luiz Alexandre Peternelli
Fabyano Fonseca e Silva

**VIÇOSA - MINAS GERAIS
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

L526m
2021
Leichtweis, Bruno Grespan, 1996-
Mapas auto-organizáveis de Kohonen para agrupamentos ambientais e estudos de interação genótipos por ambientes via normas de reação / Bruno Grespan Leichtweis. – Viçosa, MG, 2021.

44 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Felipe Lopes da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Soja - Melhoramento genético. 2. Interação genótipo-ambiente. 3. Análise de regressão. 4. Redes neurais (Neurobiologia). 5. Meio ambiente - Classificação.
I. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento.
II. Título.

CDD 22. ed. 633.342

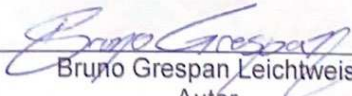
BRUNO GRESPAN LEICHTWEIS

MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN PARA AGRUPAMENTOS
AMBIENTAIS E ESTUDOS DE INTERAÇÃO GENÓTIPOS POR AMBIENTES VIA
NORMAS DE REAÇÃO

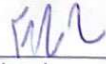
Dissertação apresentada à Universidade
Federal de Viçosa, como parte das
exigências do Programa de Pós-Graduação
em Genética e Melhoramento, para obtenção
do título de *Magister Scientiae*.

APROVADA: 23 de março de 2021

Assentimento:



Bruno Grespan Leichtweis
Autor



Felipe Lopes da Silva
Orientador

*Aos meus pais,
Jair Antonio Grespan e
Rosane Grespan Leichtweis,
Dedico.*

AGRADECIMENTOS

A Deus, a fé que me move, por encher minha vida de bênçãos, iluminar meu caminho e nunca me deixar faltar força e coragem.

Aos meus pais, pelo exemplo de caráter, por serem o chão onde me apoio, pelo incentivo e pelo esforço na realização dos meus projetos.

Ao meu orientador, Professor Felipe Lopes da Silva e coorientadores, Professor Luiz Alexandre Peternelli e Fabyano Fonseca e Silva, pela confiança e pela paciência, por todos os ensinamentos e pela amizade construída. Foi muito engrandecedor ser orientado por vocês!

Ao Curso de Pós Graduação em Genética e Melhoramento da Universidade Federal de Viçosa pela infraestrutura disponibilizada para o meu crescimento pessoal.

Aos colegas do Programa de pesquisa em melhoramento da Soja, aos colegas do laboratório de Biometria e do Biocafé, pela ajuda impagável e por todos os bons momentos que compartilhamos. Sem vocês, eu não conseguiria concluir esta etapa.

A Leticia Faria por me acompanhar nessa jornada e por toda ajuda dentro e fora da academia.

A todos os amigos e familiares que tornaram meus dias mais felizes e contribuíram para o meu crescimento pessoal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela disponibilidade da bolsa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Muito obrigado!

“Não fossem as dificuldades da vida, subestimariamos o valor da superação e não perceberíamos eminentes progressos”

Lázaro de Souza Gomes

RESUMO

LEICHTWEIS, Bruno Grespan, M.Sc., Universidade Federal de Viçosa, março de 2021. **Mapas auto-organizáveis de Kohonen para agrupamentos ambientais e estudo de interação genótipos por ambientes via normas de reação.** Orientador: Felipe Lopes da Silva. Coorientadores: Luiz Alexandre Peternelli e Fabyano Fonseca e Silva.

A cultura da soja é influenciada por diversas variáveis edáficas e climáticas. Essas variáveis representam a característica de uma região, compreender seus impactos na cultura oferece uma maior eficiência para a recomendação e implantação do material genético. A indicação de genótipos pode variar com as mudanças climáticas a nível de um macroambiente e dentro de um microambiente. Desse modo, o objetivo do trabalho foi avaliar a dissimilaridade ambiental causada por variáveis edafoclimáticas e formar estratos dentro do macriambiente sojícola três, e avaliar a interação dos genótipos pelo ambiente com base nos diferentes estratos formados mediante o estudo das normas de reação. Foram utilizados bancos de dados com informações edáficas e climáticas de 32 municípios cedidas pela empresa GDM seeds e Agrymet. Visando uma melhor caracterização dos locais em análise foi utilizada uma série histórica de avaliações dos anos de 2018 a 2020 dos meses de novembro a fevereiro. As variáveis para utilizadas para a formação dos estratos foram: registros de temperatura média ($^{\circ}\text{C}$), precipitação pluviométrica (mm), radiação solar (W.m^{-2}), velocidade do vento (m.s^{-1}), umidade relativa (%), e capacidade de armazenamento do solo (CAS). Também foram incluídas informações de latitude e altitude (m). Para o estudo da interação dos genótipos pelo ambiente, foi adotado um modelo de regressão aleatória onde os componentes de variância foram estimados pela máxima verossimilhança restrita, os valores genéticos foram previstos pelo melhor preditor linear não viesado e as normas de reação foram ajustadas através do polinômio de Legendre. A escolha da ordem do polinômio de Legendre foi selecionada pelo critério de Akaike (AIC). A metodologia do SOM possibilitou a formação de sete *clusters*, onde alguns municípios foram agrupados em *clusters* diferentes em virtude do ano de análise. O teste de diferença significativa de Tukey evidenciou diferenças entre as variáveis ambientais categóricas e contínuas, principalmente em decorrência das condições atípicas dos anos em análise. O melhor modelo de regressão aleatória foi o de ordem seis. Os genótipos S4, S15 e S20 apresentaram pouca interação com o

ambiente, enquanto os genótipos S1, S6, S9 e S17 apresentaram as maiores interações com o ambiente. As trajetórias genéticas das curvas das 20 cultivares reforçam a presença de interação genótipos por ambientes, uma vez que suas trajetórias são não-lineares e se cruzam, o que implica em uma classificação diferente a cada ambiente. Esses resultados sugerem que modelos de regressão aleatória ajustados através de polinômios de Legendre podem ser usados de forma eficiente em estudos de interação genótipos por ambientes na cultura da soja.

Palavras-chave: Regressão aleatória. Redes neurais. Classificação ambiental.

ABSTRACT

LEICHTWEIS, Bruno Grespan, M.Sc., Universidade Federal de Viçosa, March, 2021. **Kohonen's self-organizing maps for environmental groupings and genotype-environment interaction studies via reaction norms.** Adviser: Felipe Lopes da Silva. Co-advisers: Luiz Alexandre Peternelli and Fabyano Fonseca e Silva.

Soybean cultivation is influenced by several edaphic and climatic variables. These variables represent the characteristic of a region, understanding their impacts on the crop offers greater efficiency in recommending genetic material. The indication of genotypes can vary with climate change at the level of a macroenvironment and within a microenvironment. The objective of this work was to evaluate the environmental dissimilarity caused by edaphoclimatic variables in soybean macroregion three, and to evaluate the interaction of genotypes with the environment by studying the reaction norms. Databases with edaphic and climatic information from 32 municipalities provided by the company GDM seeds and Agrymet were used. Aiming at a better characterization of the places under analysis, a historical series of evaluations from the years 2018 to 2020 from the months of November to February was used. The variables used to form the strata were: records of mean temperature ($^{\circ}\text{C}$), rainfall (mm), solar radiation (Wm^{-2}), wind speed (ms^{-1}), relative humidity (%), and soil storage capacity (CAS). Latitude and altitude (m) information was also included. To study the interaction of genotypes with the environment, a random regression model was adopted where the variance components were estimated by the restricted maximum likelihood, the breeding values were predicted by the best linear unbiased predictor and the reaction norms were adjusted using the polynomial of Legendre. The choice of the order of the Legendre polynomial was selected by the Akaike criterion (AIC). The SOM methodology enabled the formation of seven clusters, where some municipalities were grouped into different clusters due to the year of analysis. Tukey's test of significant difference showed differences between categorical and continuous environmental variables, mainly due to the atypical conditions of the years under analysis. The best random regression model was the order six. Genotypes S4, S15 and S20 had little interaction with the environment, while genotypes S1, S6, S9 and S17 had the greatest interaction with the environment. The genetic trajectories of the curves of the 20 cultivars reinforce the presence of interaction between genotypes and environments, since their trajectories are non-linear and intersect, which implies a

different classification for each environment. These results suggest that random regression models fitted using Legendre polynomials can be used efficiently in studies of genotype-by-environment interaction in soybean crop.

Keywords: Random Regression. Neural Network. environment cluster.

SUMÁRIO

| | |
|--|----|
| INTRODUÇÃO GERAL | 11 |
| CAPÍTULO I | 13 |
| 1. INTRODUÇÃO..... | 16 |
| 2. MATERIAL E MÉTODOS | 17 |
| 2.1 Local de estudo..... | 17 |
| 2.2. Dados edafoclimáticos | 17 |
| 2.3 Mapa auto-organizavel de Kohonen..... | 18 |
| 3. RESULTADOS | 19 |
| 3.1. Escolha da topologia da rede neural | 19 |
| 3.2 Mapa auto-organizável de Kohonen..... | 19 |
| 3.3. Teste de comparação entre medias | 23 |
| 4. DISCUSSÃO | 24 |
| 5. CONCLUSÃO..... | 26 |
| 5. REFERÊNCIAS..... | 27 |
| CAPÍTULO II | 30 |
| 1. INTRODUÇÃO..... | 32 |
| 2. MATERIAL E MÉTODOS | 33 |
| 2.1 Material Genético..... | 33 |
| 2.2 Variáveis ambientais | 33 |
| 2.3 Componentes de variância..... | 33 |
| 3.RESULTADOS | 35 |
| 4. DISCUSSÃO..... | 39 |
| 5. CONCLUSÃO..... | 41 |
| 6. REFERENCIAS..... | 41 |

1. INTRODUÇÃO GERAL

O melhoramento de plantas tem sido conduzido por muitos anos, desde Mendel e Fisher até a aplicação atual de marcadores moleculares na seleção de genótipos superiores (BRADSHAW, 2017). No entanto, o estudo dos ambientes e a sua interação com os genótipos ($G \times E$) é uma questão importante para recomendações de genótipos superiores (BRAWNER et al., 2013; OGUT et al., 2014). A indicação de genótipos pode variar em virtude do macroambiente, com mudanças edáficas e climáticas, em diferentes latitudes, longitudes e anos (BOURRET et al., 2015; GRAY et al., 2016) e pode ainda variar com as mudanças dentro de um microambiente de um determinado local (RESENDE et al., 2016; SOARES et al., 2016).

A expressão fenotípica potencial de um dado genótipo, considerado em relação a todas as situações ambientais em que o mesmo pode sobreviver, é seu alcance de reação, ou norma de reação. Na presença de $G \times E$, a resposta diferenciada dos genótipos as variações ambientais, leva a diferentes expressões gênicas (METTLER; GREGG, 1973). A interação genótipos por ambientes ($G \times E$), por definição, refere-se ao fenômeno que ocorre quando um conjunto de genótipos tem seu desempenho alterado em ambientes diferentes.

Assim, especialmente quando se trata de caracteres com herdabilidade média a baixa, como a produtividade da soja (*Glycine max* (L.) Merrill) (DA SILVA JUNIOR et al., 2014), a interferência do meio ambiente é ainda mais problemática na expressão dessa característica (LYNCH, M., WALSH, B., 2017). A interação $G \times E$ pode reduzir a herdabilidade e ganho genético geral, a menos que os programas de melhoramento sejam estruturados para atender diferentes categorias de ambientes (LI et al., 2017; SOUSA et al., 2015). Compreender o impacto do $G \times E$ e o papel dos ambientes na sua geração, oferece oportunidades para a recomendação e implantação eficiente do material genético, otimizando os ganhos com a produtividade (OGUT et al., 2014).

Desse modo, certos genótipos são mais estáveis em diferentes ambientes. Esses são selecionados porque não apresentam comportamentos imprevisíveis e indesejados no rendimento, sendo mais resistentes às mudanças climáticas locais (EBERHART; RUSSELL, 1966). Além disso, alguns genótipos são mais adaptados, respondendo positivamente às melhorias nas condições ambientais (BRAWNER et al., 2014). A recomendação de cultivares para diferentes ambientes é eficaz quando

os genótipos podem maximizar a conversão dos recursos disponíveis para biomassa e produtividade (EL-SODA et al., 2014).

Pela importância no cenário agrícola mundial, muitos pesquisadores e empresas investem em pesquisas para compreender melhor os fatores ambientais que contribuem para a maximização da produtividade. O Brasil tem proporções continentais, estima-se que existam atualmente 36,8 milhões hectares de área plantada no território nacional (EMBRAPA SOJA, 2020). Vários estudos (ALLIPRANDINI et al., 2009; SOUSA et al., 2015; XAVIER et al., 2018) demonstraram ganhos de produtividade de acordo com a disponibilidade de luz, água e nutrientes.

Embora os autores mencionados abordem o uso de diferentes métodos no estudo da interação $G \times E$, não há estudos na literatura que examinem uma recomendação de acordo com a mudança ambiental contínua na região. Normalmente as variáveis ambientais são atribuídas como fenômenos discretos, gerando grupos com características ambientais semelhantes, de modo que os ambientes são tratados como níveis de variáveis categóricas (COOPER; DELACY, 1994). Considerar o ambiente como uma variável categórica significa perder a capacidade para diagnosticar mudanças suaves e graduais do ambiente (EL-SODA et al., 2014).

Em uma abordagem de agrupamento usando redes neurais artificiais (RNAs), como os mapas auto-organizáveis de Kohonen, as informações ambientais podem ser representadas por grupos. Porém, com uso de distância entre os grupos, as variáveis podem ser representadas de forma espacial contínua (KOHONEN, 2013). Portanto, o uso de RNAs pode ser considerado uma ferramenta útil na caracterização edáfica e climática de uma região. Associando essas informações a modelos de regressão aleatória, que possuem a vantagem de quantificar as respostas do genótipo às mudanças no ambiente de forma linear (KOLMODIN; BIJMA, 2004), podemos analisar a interação de características que variam continuamente ao longo de um ambiente com a expressão dos diferentes genótipos, aumentando a eficiência na recomendação de novas cultivares.

Tendo em vista o exposto acima, esse trabalho teve por objetivo: avaliar a dissimilaridade ambiental causada por variáveis edafoclimáticas e formar estratos ambientais, utilizando as metodologias de Mapas Auto-organizáveis de Kohonen; e, propor um método para otimizar recomendações de genótipos de acordo com gradientes ambientais, com o uso de informações edáficas e climáticas.

CAPÍTULO I

USO DE MAPAS AUTO-ORGANIZAVEIS PARA ESTUDO DE DIVERSIDADE AMBIENTAL

RESUMO

LEICHTWEIS, Bruno Grespan, M.Sc., UNIVERSIDADE FEDERAL DE VIÇOSA, março de 2021. **Uso de mapas auto-organizáveis para estudo de diversidade ambiental.** Orientador: Felipe Lopes da Silva. Coorientadores: Luiz Alexandre Peternelli e Fabyano Fonseca e Silva

A cultura da soja é influenciada por diversas variáveis edáficas e climáticas, essas variáveis representam as características de uma região, assim se faz necessário resumir e conhecer essas informações para uma melhor recomendação e implantação do material genético, otimizando os ganhos com a produtividade. Dentre as metodologias utilizadas para se classificar essas regiões, o uso de redes neurais em particular o mapa de auto-organização (SOM) se destaca por classificar de forma eficaz a associação desses genótipos em diferentes ambientes. Desse modo, o objetivo foi avaliar a dissimilaridade causada por variáveis edafoclimáticas e formar estratos dentro do macroambiente sojícola três, utilizando as metodologias de Mapas Auto-organizáveis de Kohonen. Foram utilizados bancos de dados com informações edáficas e climáticas de 32 municípios cedidas pela empresa GDM seeds e Agrymet. Visando uma melhor caracterização dos locais em análise foi utilizada uma série histórica de avaliações dos anos de 2018 a 2020 dos meses de novembro, dezembro, janeiro e fevereiro utilizando-se registros de temperatura média ($^{\circ}\text{C}$), precipitação pluviométrica (mm), radiação solar (W.m^{-2}), velocidade do vento (m.s^{-1}), umidade relativa (%), capacidade de armazenamento do solo (CAS). Também foram incluídas informações de latitude e altitude (m). A fim de se caracterizar os padrões distribuição ambiental dos municípios em estudo foi realizada a metodologia de mapa de auto-organização (SOM) seguida de uma análise de *cluster* hierárquica segundo o método de ligação de Ward usando a distância euclidiana para definir os limites do *cluster* nas unidades do SOM e para comparar as diferenças entre os clusters foi realizado o teste de diferença significativa de Tukey (HSD). Dentre as topologias de rede testadas, a medida de dissimilaridade do quadrado médio da distância euclidiana (Modelo 4) apresentou o menor valor de erro topográfico (0,186) e, portanto, foi utilizada para a realização da metodologia de SOM. A metodologia possibilitou a formação de sete *clusters*, onde alguns municípios foram agrupados em *clusters* diferentes em virtude do ano de análise. O teste de diferença significativa de Tukey evidenciou diferenças entre as variáveis categóricas e contínuas, principalmente em decorrência das condições atípicas dos anos em análise. A metodologia SOM se mostrou altamente

eficiente para um melhor entendimento de ambientes dinâmicos captando uma maior sensibilidade a mudanças nas variáveis observadas, como no caso da macrorregião sojícola 3. No entanto, uma vez que a abordagem de redes neurais não foi projetada para a avaliação de interações com genótipos que seriam mais indicados para essa região, os resultados obtidos podem ser usados em combinação com modelos de regressão aleatória.

Palavras-chave: Redes neurais; Dissimilaridade ambiental; Características edafoclimáticas

1. INTRODUÇÃO

A cultura da soja é influenciada por diversas variáveis edáficas e climáticas, e são essas variáveis que representam as características de uma região geográfica (SOUSA et al., 2015). As informações edáficas e climáticas podem ser afetadas por muitos fatores como o grande volume de dados, por apresentarem estruturas complexas, não-lineares que exibem ruído, redundância, relacionamentos internos e *outliers* (GIANOLA et al., 2011), além da existência de interações complexas existentes entre as variáveis explicativas e de resposta. Assim, metodologias não-lineares são preferíveis para lidar com esse conjunto de dados ambientais complexos (CALUS; BIJMA; VEERKAMP, 2004; GIANOLA et al., 2011).

Desse modo, a metodologia não-linear com o uso de redes neurais artificiais, tornou-se cada vez mais popular nos últimos anos. Dentre essas metodologias, o mapa auto-organizável (SOM) surgiu como uma ferramenta versátil para extrair informações a partir de dados complexos podendo ser aplicado de forma eficaz na classificação e associação de amostras e suas variáveis (KOHONEN, 2013).

As redes neurais de Kohonen ou mapas auto organizáveis (SOM) são modelos computacionais inicialmente desenvolvidos pelo professor Teuvo Kohonen (KOHONEN, 1990) e são utilizados principalmente para análises de agrupamentos. As redes SOM possuem aprendizado competitivo e não supervisionado, em que a base dos algoritmos de adaptação dos pesos sinápticos é a competição global entre os neurônios. Permitindo, a construção de um mapa topológico a partir de uma rede, em que neurônios que estão topologicamente próximos respondam, de forma semelhante, a padrões de entrada semelhantes (HAYKIN, 2017).

Nesse contexto, os programas de melhoramento da soja visam, dentre outros objetivos, a seleção de genótipos com desempenho superior (REZENDE; CARVALHO, 2007). A seleção desses genótipos, além de ser pautada nas condições climáticas e edáficas discutidas acima, deve se atentar às interações do genótipo com o ambiente (G x E). A escolha correta da região onde esses cultivares serão desenvolvidos pode ser decisiva no lançamento de cultivares mais competitivas no mercado (SQUILASSI, 2003).

Nesse contexto, os ensaios multi-ambientais (METs) são amplamente usados pelos melhoristas de plantas para avaliar o desempenho relativo dos genótipos para uma região-alvo de ambientes. A análise estatística dos dados obtidos a partir desses

METs tem recebido muita atenção, principalmente em resposta às dificuldades geradas pelas interações G × E (DING et al., 2008). Conseqüentemente, diversas metodologias têm sido propostas, desenvolvidas e utilizadas para essas análises.

Diante do exposto, objetivou-se avaliar a dissimilaridade ambiental causada por variáveis edafoclimáticas e formar estratos ambientais, utilizando as metodologias de Mapas Auto-organizáveis de Kohonen.

2. MATERIAL E MÉTODOS

2.1 Local de estudo

Os dados climáticos em análise são resultados de avaliações em 32 municípios (Tabela 2) situadas dentro da macrorregião sojícola brasileira denominada de M3 (MRS 3), nos estados de Goiás e de Minas Gerais. Essa região agrícola possui como características principais o clima subtropical úmido com verões quentes e chuvosos, invernos secos e de temperaturas amenas (KASTER; FARIAS, 2002). A escolha dessa região se deve principalmente pela forte interação entre genótipos e o ambiente, necessitando de precauções no cultivo e recomendação de genótipos de soja (EMBRAPA SOJA, 2020).

Os municípios escolhidos fazem parte da rede de ensaios de valor de cultivo e uso (VCU) da empresa GDM Genética Brasil (GDM). Estes ensaios caracterizam-se pela última fase de seleção, quando os genótipos são avaliados em vários locais (ensaios regionais) para a determinação de adaptabilidade e estabilidade fenotípicas.

2.2. Dados edafoclimáticos

A fim de se caracterizar os ambientes de cada ensaio, foram utilizados bancos de dados com informações edáficas e climáticas gentilmente cedidas pela empresa Agrymet em parceria com a GDM. Visando uma melhor caracterização dos locais em análise foi utilizada uma série histórica de avaliações dos anos de 2018 a 2020 dos meses de novembro, dezembro, janeiro e fevereiro.

Foram analisados registros de temperatura média (°C), precipitação pluviométrica (mm), radiação solar ($W.m^{-2}$), velocidade do vento ($m.s^{-1}$), umidade relativa (%), e capacidade de armazenamento do solo (CAS). Também foram incluídas informações de latitude e altitude (m).

2.3 Mapa auto-organizável de Kohonen

A metodologia de mapa de auto-organização (SOM) foi usada para caracterizar os padrões de distribuição espacial das variabilidades ambientais nos 32 municípios em estudo. O SOM aproxima a função de densidade de probabilidade dos dados de entrada e pode ser usado para agrupamento e visualização eficiente de dados de alta dimensão em uma rede bidimensional (KOHONEN, 2013).

Como dados de entrada foram utilizadas as variáveis edafoclimáticas. No presente estudo, treinamos o SOM com diferentes tamanhos de mapas e selecionamos como o número apropriado de neurônios de saída com base na equação heurística $(5\sqrt{n})$ (numero de amostras) (VESANTO, J., HIMBERG, J., ALHONIEMI, E., AND PARHANKANGAS, 2000).

Durante o processo de aprendizagem do SOM, cada vetor de entrada foi atribuído a um neurônio de saída por meio do cálculo do SOM. Cada neurônio de saída apresentou um vetor de coeficiente ou pesos associados aos dados de entrada; esses pesos estabeleceram uma ligação entre as unidades de entrada e suas unidades de saída associadas (RITTER; KOHONEN, 1989).

Para medir o desempenho quantitativo dos diferentes modelos testados, e estabelecer um nível apropriado de confiança, foi utilizado o critério de erro topográfico. Esse critério foi adotado por ser indicador da precisão do mapeamento e na preservação da topologia (KOHONEN, 2013).

Após o processo de aprendizagem do SOM, foi realizada uma análise de *cluster* hierárquica segundo o método da variância mínima de Ward para os neurônios usando a distância euclidiana para definir os limites do *cluster* nas unidades do SOM. Além disso, a análise de variância (ANOVA) e o teste de diferença significativa de Tukey (HSD) foram realizados para comparar as diferenças nas variáveis entre os clusters definidos no SOM. Os elementos dos vetores de peso para diferentes variáveis de entrada foram visualizados para caracterizar sua contribuição em relação aos *clusters* no SOM treinado em escala de cores. Todas as análises foram realizadas no software R (versão 3.6.0). Para a construção das redes utilizou-se o pacote “Kohonen” (WEHRENS; BUYDENS, 2007).

3. RESULTADOS

3.1. Escolha da topologia da rede neural

Na tabela 1 estão representados os resultados dos modelos testados com diferentes topologias de rede e medidas de dissimilaridade e seus respectivos erros topográficos. Os erros topográficos variaram de 0,186 (Modelo 4) a 0,835 (Modelo 6). Dos oito modelos testados para o conjunto de dados em análise, o modelo que apresentou o menor valor de erro topográfico (0,186) e conseqüentemente se mostrou o modelo mais adequado foi o modelo 4. Ao se analisar as medidas de dissimilaridade, foi possível identificar que o quadrado médio da distância euclidiana apresentou os menores valores de erro topográfico nas duas topologias de rede testadas (11 x 8 e 10 x 9).

Tabela 1: Modelos testados com diferentes medidas de dissimilaridade, tamanho de mapa e erro topográfico para agrupamento dos 32 municípios em estudo.

| Modelos | Medida de dissimilaridade | Tamanho do mapa | Erro topográfico |
|---------|---------------------------|-----------------|------------------|
| 1 | Euclidiana | 88 (11 x 8) | 0,352 |
| 2 | Manhattan | 88 (11 x 8) | 0,814 |
| 3 | Tanimoto | 88 (11 x 8) | 0,692 |
| 4 | QMDE | 88 (11 x 8) | 0,186 |
| 5 | Euclidiana | 90 (10 x 9) | 0,361 |
| 6 | Manhattan | 90 (10 x 9) | 0,835 |
| 7 | Tanimoto | 90 (10 x 9) | 0,687 |
| 8 | QMDE | 90 (10 x 9) | 0,218 |

QMDE: Quadrado médio da distância Euclidiana.

3.2. Mapa auto-organizável de Kohonen

A representação gráfica dos grupos formados é mostrada na figura 1A. Um corte feito à distância 1,45, que corresponde a 94% do valor da distância no último nível de fusão (24,29), possibilitou estabelecer sete grupos. A figura apresenta ainda um *heatmap* onde se encontram representados a distribuição dos pesos de cada variável em relação aos 88 neurônios formados com a metodologia de mapas auto-organizáveis de Kohonen, baseado no modelo 4 selecionado (Tabela 1).

A escala de cores apresentado na figura 1A representa os pesos sinápticos dados a cada uma das variáveis. Essa escala de cores, varia de cores azuis, com menores valores de pesos para cada variável; a cores amarelas, com maiores pesos

sinápticos para cada variável. Ao se avaliar a formação dos grupos, fica clara a importância de cada variável na formação desses, como a diferenciação do primeiro grupo pelo dendograma, apenas as variáveis altitude e latitude não foram importantes para discriminar a sua formação.

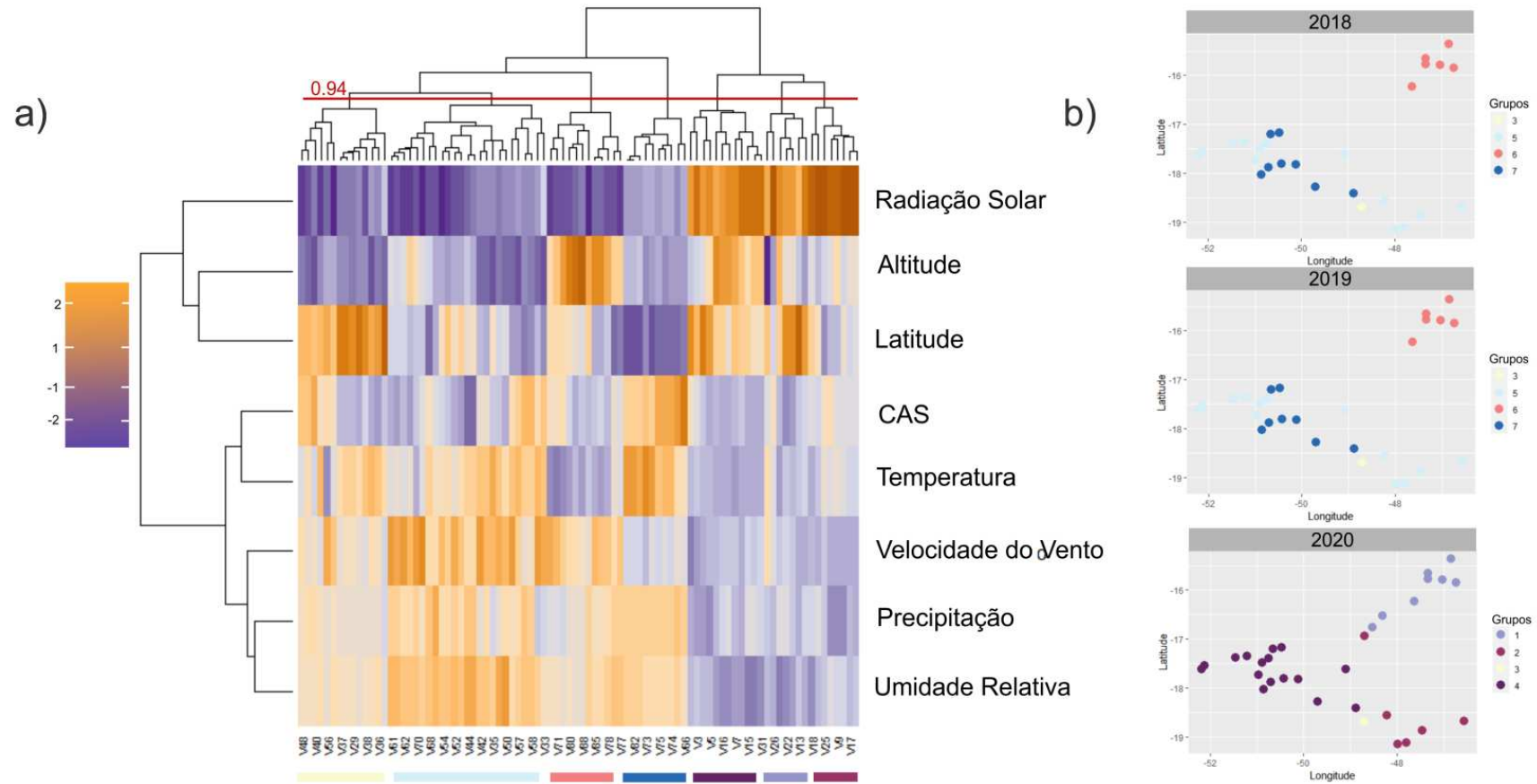


Figura 1: a: *Heatmap* e dendrograma formados a partir da metodologia de mapas auto-organizáveis de Kohonen utilizando informações edafoclimáticas dos locais em estudo. A variação da escala de cores no *heatmap* representa o peso sináptico dos 88 neurônios para cada variável. As cores abaixo do *heatmap* representam a classificação dos neurônios dentro de cada grupos. b: Disposição geográfica dos municípios pertencentes aos sete grupos formados pelo método de agrupamento de Ward. CAS: Capacidade de Armazenamento do Solo. V_n são as chamadas dos 88 neurônios.

Na figura 1B se encontram as descrições dos *clusters* dos municípios em análise nos anos de avaliação. Os grupos de 1 a 7 estão representados por diferentes cores e os municípios que compõem cada grupo se encontram na Tabela 2.

Tabela 2. Municípios utilizados na metodologia de mapas auto-organizáveis de kohonen, com seus respectivos estados e o resultado da classificação dos grupos nos três anos de avaliação.

| Municípios | Estado | Grupo pertencente por ano | | |
|----------------------------|--------|---------------------------|------|------|
| | | 2018 | 2019 | 2020 |
| Acreúna | GO | 7 | 7 | 4 |
| Araguari | MG | 5 | 5 | 2 |
| Bom Jesus | GO | 7 | 7 | 4 |
| Buritís | MG | 6 | 6 | 1 |
| Cabeceiras | GO | 6 | 6 | 1 |
| Cristalina | GO | 6 | 6 | 1 |
| Jataí | GO | 7 | 7 | 4 |
| Luziânia | GO | NA | NA | 1 |
| Mineiros | GO | 5 | 5 | 4 |
| Monte Carmelo | MG | 5 | 5 | 2 |
| Montevídiu 1 | GO | 7 | 7 | 4 |
| Montevídiu 2 | GO | 5 | 5 | 4 |
| Morrinhos | GO | 5 | 5 | 4 |
| Nova Ponte | MG | 5 | 5 | 2 |
| Paraúna | GO | 5 | 5 | 4 |
| Patos de Minas | MG | 5 | 5 | 2 |
| Perolândia | GO | 5 | 5 | 4 |
| Planaltina 1 | GO | 6 | 6 | 1 |
| Planaltina 2 | GO | 6 | 6 | 1 |
| Planaltina 3 | GO | 6 | 6 | 1 |
| Planalto Verde | GO | 5 | 5 | 4 |
| Rio Verde 1 | GO | 5 | 5 | 4 |
| Rio Verde 2 | GO | 5 | 5 | 4 |
| Santa Helena de Goiás | GO | 7 | 7 | 4 |
| São Miguel do Passa Quatro | GO | NA | NA | 2 |
| Serranópolis | GO | 7 | 7 | 4 |
| Silvania | GO | NA | NA | 1 |
| Tupaciguara 1 | MG | 7 | 7 | 4 |
| Tupaciguara 2 | MG | 3 | 3 | 3 |
| Turvelândia | GO | 7 | 7 | 4 |
| Uberlândia | MG | 5 | 5 | 2 |
| Unai | MG | 6 | 6 | 1 |

* NA representa a ausência de informação para a análise de agrupamento em anos específicos.

Ao se analisar o comportamento dos grupos ao longo dos anos de avaliação, foi possível concluir que os municípios pertencentes aos grupos 6 e 7 tiveram comportamentos climáticos diferentes no ano de 2020, sendo classificados como 1 e 4 respectivamente. Enquanto os municípios pertencentes ao grupo 5, no ano de 2020 foram classificadas em dois grupos, o 2 e 4. O agrupamento 3 não teve alteração de classificação nos diferentes anos de avaliação.

A utilização do método de agrupamentos de mapas auto-organizáveis de kohonen, utilizando a distância QME apresentou alto poder discriminatório para esse estudo de diversidade ambiental, uma vez que foram utilizados dados de informações contínuas (variáveis climáticas) e categóricas (variáveis edáficas).

3.3. Teste de comparação entre médias

Os resultados dos testes para as diferentes variáveis em estudo comparando os agrupamentos formados pelo método hierárquico de Ward estão apresentados na Figura 2. Para todas as variáveis foram encontradas diferenças significativas entre as médias dos grupos ($p < 0.05$).

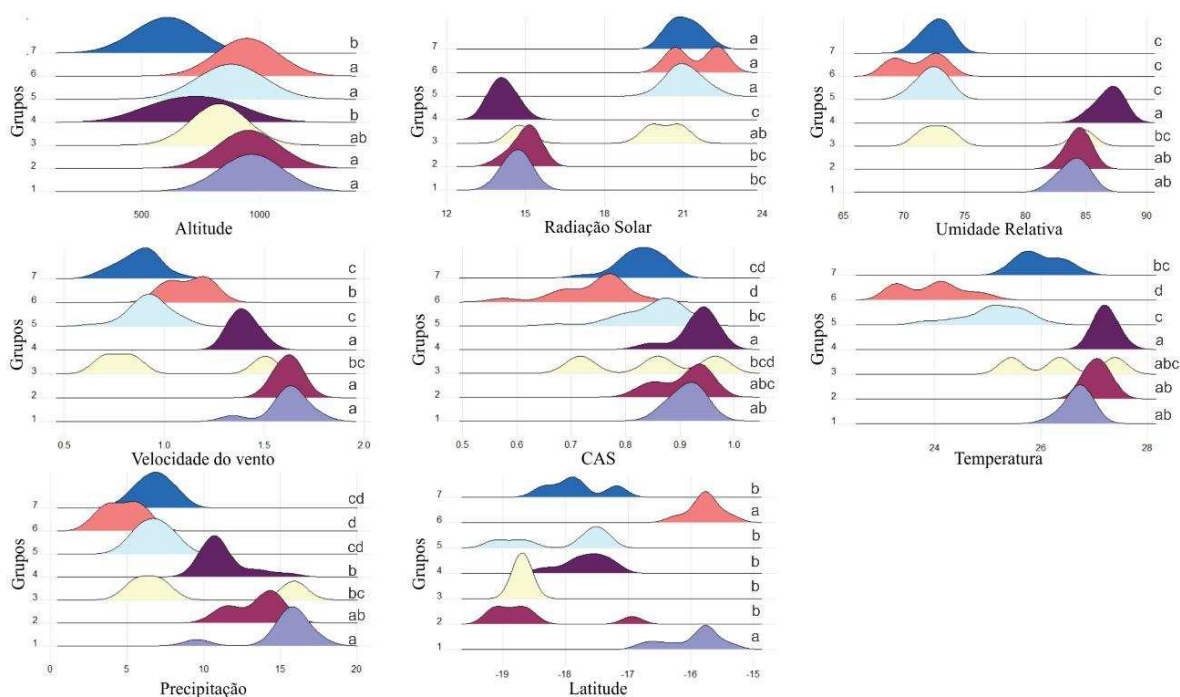


Figura 2. Diferenças entre médias aplicadas pelo teste de Tukey ($p = 0,05$) entre grupos das diferentes variáveis em estudo. Médias seguidas por letras diferentes na mesma coluna diferiram entre si pelo teste de Tukey a 5% de significância.

Para a variável altitude era esperado a igualdade das médias entre os grupos 6 e 1, 7 e 4, pois estes representam os mesmos municípios, em anos de avaliação diferentes. As diferenças foram entre os grupos com médias mais altas (1,2,5 e 6), com as mais baixas (7 e 4).

Os anos de 2018 e 2019 foram classificados com médias menores nos grupos 5,6 e 7 para as variáveis umidade relativa, temperatura, precipitação e velocidade do vento e com alta radiação solar, comparado com os mesmos municípios que formaram os grupos 1,2 e 4, representando o ano de 2020. Para este ano, os grupos tiveram baixas médias de radiação solar e altas medias de umidade relativa, temperatura, precipitação e velocidade do vento.

O grupo 3, composto por apenas uma cidade, apresentou diferentes distribuições dos dados para todas as variáveis contínuas, esse fato pode ser explicado pelo grupo ser o único que apresenta informações dos três anos de avaliação. Para as variáveis categóricas, o grupo não apresentou diferença entre as médias apenas para a altitude. Para as variáveis contínuas, este se apresentou diferente do grupo 4 em umidade relativa e dos grupos 1, 2 e 4 em velocidade do vento. O mesmo apresenta ter maior média de temperatura que o grupo 6 e ter a pluviosidade entre os grupos 1 e 6.

4. DISCUSSÃO

O entendimento sobre as condições edáficas e climáticas de uma região, são de suma importância para o melhoramento de qualquer cultura, uma vez que certos genótipos são mais estáveis em diferentes ambientes, esses materiais são selecionados porque não apresentam modificações indesejadas no rendimento, sendo mais resiliente às mudanças climáticas locais (EBERHART; RUSSELL, 1966). Além disso, alguns genótipos são mais adaptativos, respondendo positivamente às melhorias nas condições ambientais (BRAWNER et al., 2014).

No presente estudo, buscou-se classificar e analisar, através da metodologia de mapas auto-organizáveis, uma serie temporal de dados sob cenário de mudança dinâmica do clima na macrorregião sojícola brasileira 3. O uso de redes neurais artificiais (RNAs) se mostrou altamente eficiente, pois a forma como estes são apresentados internamente pela rede cria a possibilidade de um desempenho superior ao dos modelos convencionais (HAYKIN, 2017). Esse modelo foi sensível às

mudanças do macroambiente 3, e a escolha do método de agrupamento de Ward evidenciou a homogeneidade dos sete grupos formados não sofrendo um viés em virtude dos *outliers* presentes nos dados em análise.

Em contraste com as abordagens convencionais, o presente estudo buscou uma abordagem sensível ao ambiente dinâmico. Portanto, na formulação de grupos ambientais, as informações de topografia, como altitude e latitude são importantes para uma interpretação inicial da rede, enquanto as mudanças subsequentes na formação dos grupos são assumidas como sendo impulsionadas pela dinâmica das variáveis edafoclimáticas contínuas (temperatura, umidade relativa, radiação solar, CAS, velocidade do vento e precipitação) ao longo do tempo.

Os resultados do cenário de mudança climática entre anos de avaliação, com base no modelo do SOM, sugerem que a resposta dos agrupamentos difere distintamente entre os locais. A macrorregião sojícola brasileira 3 é caracterizada com baixas umidades relativas e altos valores de radiação solar, com a região norte se destacando por temperaturas menores comparadas às outras regiões (KASTER; FARIAS, 2002). Ao analisar as variações dos agrupamentos ao longo dos diferentes anos de avaliação, verificou-se uma menor coincidência de agrupamento no ano 2020 quando comparado aos demais. No ano de 2020 foi caracterizado pela maior média de temperatura ambiental com precipitação maiores na região norte.

Embora existam diferentes abordagens no estudo da interação $G \times E$, ainda não há estudos na literatura que descrevem uma recomendação de acordo com a mudança ambiental contínua. Normalmente as variáveis ambientais são atribuídas como fenômenos discretos, gerando grupos com características ambientais semelhantes, de modo que os ambientes são tratados como níveis de variáveis categóricas (COOPER; DELACY, 1994). A modelagem da variação espacial e da dinâmica temporal é um desafio para estudos de interação na cultura da soja.

Uma vez que a abordagem das redes neurais não foi projetada para avaliar as interações, os resultados obtidos podem ser usados em combinação com modelos de regressão aleatória: como uma primeira etapa, a dinâmica ambiental é estudada a partir de um modelo de mapas auto organizáveis de kohonen e pode então, como próxima etapa, ser analisada a interação genótipo-ambiente com base em modelos de regressão aleatória.

5. CONCLUSÃO

Neste estudo, foi apresentado um modelo de mapas auto-organizáveis para modelagem de dados de séries temporais para classificação ambiental sob o cenário de mudança dinâmica do clima na macrorregião sojícola três. A utilização do método de agrupamentos de mapas auto-organizáveis de kohonen, utilizando a distância QME, foi eficiente na formação de estratos ambientais utilizando informações edafoclimáticas contínuas e categóricas.

6. REFERÊNCIAS

- ALLIPRANDINI, L. F. et al. Understanding soybean maturity groups in brazil: Environment, cultivar classification, and stability. **Crop Science**, 2009.
- ALVES, R. S. et al. Optimization of Eucalyptus breeding through random regression models allowing for reaction norms in response to environmental gradients. **Tree Genetics & Genomes**, 2020.
- BOURRET, A. et al. Multidimensional environmental influences on timing of breeding in a tree swallow population facing climate change. **Evolutionary Applications**, 2015.
- BRADSHAW, J. E. Plant breeding: past, present and future. **Euphytica**, v. 213, n. 3, p. 1–12, 2017.
- BRAWNER, J. T. et al. Classifying genotype by environment interactions for targeted germplasm deployment with a focus on Eucalyptus. **Euphytica**, 2013.
- BRAWNER, J. T. et al. Visualising the environmental preferences of pinus tecunumanii populations. **Tree Genetics and Genomes**, 2014.
- CALUS, M. P. L.; BIJMA, P.; VEERKAMP, R. F. Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model. **Genetics Selection Evolution**, 2004.
- COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, 1994.
- DA SILVA JUNIOR, C. A. et al. Correlations and genetic parameters between morphological descriptors in soybean. **Journal of Agronomy**, 2014.
- DING, M. et al. **Multi-environment trial analysis for Pinus radiata**. New Zealand Journal of Forestry Science. **Anais...**2008
- DUARTE, J. B.; VENCOSKY, R. Estimaco e predico por modelo linear misto com ênfase na ordenaco de mdias de tratamentos genticos. **Scientia Agricola**, 2001.
- EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties 1. **Crop Science**, 1966.
- EL-SODA, M. et al. **Genotype x environment interaction QTL mapping in plants: Lessons from Arabidopsis**Trends in Plant Science, 2014.
- EMBRAPA SOJA. **Tecnologias de Produo de Soja - Regio Central do Brasil 2020**. [s.l: s.n.].
- GIANOLA, D. et al. Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. **BMC Genetics**, 2011.
- GILMOUR, A R. et al. ASReml User Guide Release 4.1 Structural Specification. **VSN International Ltd. Hemel Hempstead**, 2015.
- GOMULKIEWICZ, R.; KIRKPATRICK, M. Quantitative genetics and the evolution of reaction norms. **Evolution**, 1992.

GRAY, L. K. et al. Developing management strategies for tree improvement programs under climate change: Insights gained from long-term field trials with lodgepole pine. **Forest Ecology and Management**, 2016.

HAYKIN, S. Redes neurais: princípios e prática. **Bookman**, 2017.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, 1975.

KASTER, M.; FARIAS, J. R. B. **REGIONALIZAÇÃO DOS TESTES DE VALOR DE CULTIVO E USO E DA INDICAÇÃO DE CULTIVARES DE SOJA**. XXIV Reunião de Pesquisa de Soja da Região Central do Brasil. **Anais...**2002

KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics**, 1990.

KOHONEN, T. The Self-Organizing Map. **Proceedings of the IEEE**, 1990.

KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, 2013.

KOLMODIN, R.; BIJMA, P. Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. **Genetics Selection Evolution**, 2004.

LANDE, R.; ARNOLD, S. J. The Measurement of Selection on Correlated Characters. **Evolution**, 1983.

LI, Y. et al. **Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application** **Tree Genetics and Genomes**, 2017.

LYNCH, M., WALSH, B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA, USA. **CEUR Workshop Proceedings**, 2017.

METTLER, L. E.; GREGG, T. G. Population genetics and evolution. In: São Paulo, Poligono, Ed. da USP: [s.n.]. p. 262 p.

MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. **Genetics Selection Evolution**, 1998.

OGUT, F. et al. Genetic analysis of Diallel progeny test data using factor analytic linear mixed models. **Forest Science**, 2014.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, 1971.

R DEVELOPMENT CORE TEAM, R. **R: A Language and Environment for Statistical Computing**. [s.l: s.n.].

RAYMOND, C. A. Genotype by environment interactions for Pinus radiata in New South Wales, Australia. **Tree Genetics and Genomes**, 2011.

RESENDE, R. T. et al. Intra-genotypic competition of Eucalyptus clones generated by environmental heterogeneity can optimize productivity in forest stands. **Forest**

Ecology and Management, 2016.

REZENDE, P. M. DE; CARVALHO, E. DE A. Avaliação de cultivares de soja [*Glycine max* (L.) Merrill] para o sul de Minas Gerais. **Ciência e Agrotecnologia**, v. 31, n. 6, p. 1616–1623, 2007.

RITTER, H.; KOHONEN, T. Self-organizing semantic maps. **Biological Cybernetics**, 1989.

SCHAEFFER, L. R. Application of random regression models in animal breeding. **Livestock Production Science**, 2004.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, 1978.

SOARES, A. A. V. et al. Increasing stand structural heterogeneity reduces productivity in Brazilian Eucalyptus monoclonal stands. **Forest Ecology and Management**, 2016.

SOUSA, L. B. et al. Evaluation of soybean lines and environmental stratification using the AMMI, GGE biplot, and factor analysis methods. **Genetics and Molecular Research**, 2015.

SQUILASSI, M. G. **Interação de genótipos com ambientes**. [s.l: s.n.].

VESANTO, J., HIMBERG, J., ALHONIEMI, E., AND PARHANKANGAS, J. SOM Toolbox for Matlab 5, Report A57. **Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland**, 2000.

WEHRENS, R.; BUYDENS, L. M. C. Self- and super-organizing maps in R: The kohonen package. **Journal of Statistical Software**, 2007.

WOLTERECK, R. Weitere experimentelle untersuchungen über Artänderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden. **Zeitschrift für Induktive Abstammungs- und Vererbungslehre**, 1913.

XAVIER, A. et al. Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. **G3: Genes, Genomes, Genetics**, 2018.

CAPÍTULO II

NORMAS DE REAÇÃO PARA ESTUDOS DE INTERAÇÃO GENÓTIPOS POR AMBIENTES E INDICAÇÃO DE CULTIVARES SUPERIORES DE SOJA

RESUMO

LEICHTWEIS, Bruno Grespan, M.Sc., Universidade Federal de Viçosa, Março de 2021. **Normas de reação para estudos de interação genótipos ambientes e indicações de cultivares superiores de soja.** Orientador: Felipe Lopes da Silva. Coorientadores: Luiz Alexandre Peternelli e Fabyano Fonseca e Silva.

A cultura da soja é influenciada por diversas variáveis edáficas e climáticas, essas variáveis representam uma região edafoclimática. Compreender os impactos dessas variáveis oferece uma maior eficiência para a recomendação e implantação do material genético, otimizando os ganhos com a produtividade. Nesse contexto, a utilização da metodologia de norma de reação apresenta como vantagens a previsão na expressão fenotípica do genótipo em qualquer ambiente, e também quantifica a sensibilidade no traço do ambiente através da inclinação de uma norma de reação linear. Assim, objetivou-se avaliar a interação genótipo-ambiente para a característica produtividade mediante o estudo de normas de reação ao ambiente, obtidas pelo modelo de regressão aleatória, e selecionar os genótipos mais adaptados com base em variações ambientais contínuas. Para isso, a característica produtividade de grãos de 20 genótipos foi avaliada por três anos. Os componentes de variância foram estimados pela máxima verossimilhança restrita (REML), os valores genéticos foram previstos pelo melhor preditor linear não viesado (BLUP) e as normas de reação foram ajustadas através do polinômio de Legendre. A escolha da ordem do polinômio de Legendre foi selecionada pelo critério de Akaike (AIC). De acordo com o critério de AIC, o melhor modelo de regressão aleatória foi o de ordem seis. Os genótipos S4, S15 e S20 apresentaram pouca interação com o ambiente, enquanto os genótipos S1, S6, S9 e S17 apresentaram as maiores interações com o ambiente. As trajetórias genéticas das curvas das 20 cultivares de soja reforçam a presença de interação genótipo x ambiente, uma vez que suas trajetórias são não-lineares e se cruzam, o que implica em uma classificação diferente a cada ambiente. Esses resultados sugerem que modelos de regressão aleatória ajustados através de polinômios de Legendre podem ser usados de forma eficiente em estudos de interação genótipo-ambiente na cultura da Soja.

Palavras-chave: Regressão Aleatória; Modelos Mistos; Polinômio de Legendre.

1. INTRODUÇÃO

A interação genótipo por ambiente ($G \times E$) pode ser definida como o fenômeno que ocorre quando um conjunto de genótipos tem seu desempenho alterado em diferentes ambientes. Na presença de $G \times E$, os genótipos podem apresentar respostas diferenciadas devido às variações ambientais, levando a diferentes expressões gênicas (LI et al., 2017; SOUSA et al., 2015). Em um programa de melhoramento, costuma-se selecionar genótipos estáveis, aqueles que não apresentam sensibilidade para mudanças ambientais, porém, o nível da expressão pode ser grande aliada para decidir qual material genético deve ser selecionado para implantação em ambientes de cultivo específicos, maximizando o desempenho genético local (RAYMOND, 2011).

A interação $G \times E$ pode ser categorizada em duas classes: (1) interação causando mudança na classificação, em que os genótipos são classificados em diferentes ordens em diferentes ambientes; e (2) interação causando diferentes níveis de expressão do mesmo genótipo em diferentes ambientes (normas de reação), não necessariamente havendo mudança na ordem da classificação dos genótipos. Cabe ao melhorista, portanto, determinar e entender as relações e magnitude da interação $G \times E$ para se obter um melhor desempenho genético (RAYMOND, 2011).

Na cultura da soja, uma série de metodologias são utilizadas para medir a extensão de $G \times E$ para características no melhoramento. Os métodos tradicionais são baseados na análise de variância (ANOVA) e regressão linear. No entanto, esses métodos apresentam limitações pela falta de capacidade em explicar grande parte da interação, não informando sobre interações específicas positivas ou negativas com os ambientes (DUARTE; VENCOSKY, 2001).

Nesse contexto, a metodologia de norma de reação descreve uma gama de respostas ou fenótipos produzidos por um único genótipo em um intervalo de ambientes (LYNCH, M., WALSH, B., 2017; WOLTERECK, 1913). Esta metodologia é adequada para analisar dados em características que variam gradualmente e continuamente ao longo de um gradiente ambiental. A vantagem da norma de reação é que a resposta de seleção pode ser prevista não apenas na expressão fenotípica em qualquer ambiente, mas também na quantificação da sensibilidade no traço do ambiente através da inclinação de uma norma de reação linear (KOLMODIN; BIJMA, 2004).

A norma de reação pode ser explicada como a expressão fenotípica potencial de um dado genótipo, considerado em relação a todas as situações ambientais em que o genótipo pode sobreviver. Quando as normas são relativamente estreitas, significa que a expressão fenotípica é previsivelmente constante. Por outro lado, se o alcance das reações são grandes, o desenvolvimento é mais flexível e diversos fenótipos são produzidos sob condições ambientais adversas (METTLER; GREGG, 1973).

Diante do exposto, objetivou-se avaliar a interação genótipos-ambientes para a característica produtividade de grãos de soja, mediante o estudo de normas de reação ao ambiente, obtidas pelo modelo de regressão aleatória, e selecionar os genótipos mais adaptados com base em variações ambientais contínuas.

2. MATERIAL E MÉTODOS

2.1. Material genético

O conjunto de dados utilizados nesse estudo foram cedidos gentilmente pela empresa GDM Genética do Brasil. Trata-se de resultados de produtividade de grãos de soja, em kg/ha, determinados em ensaios de valor de cultivo e uso (VCU) em que 20 genótipos foram avaliados. Esses ensaios caracterizam-se pela última fase de seleção, quando os genótipos são avaliados em vários locais (ensaios regionais) para a determinação de adaptabilidade e estabilidade fenotípicas.

2.2 Variáveis ambientais

As variáveis ambientais foram obtidas através do agrupamento realizado pelo método de mapas auto-organizáveis de Kohonen de variáveis edáficas e climáticas em três anos de avaliação, como apresentadas no capítulo 1, formando sete grupos ambientais. A ordem dos grupos foi obtida pela média de produtividade dos genótipos classificados dentro de cada extrato e, ordenados do grupo menos produtivo ao mais produtivo. Seguindo essa ordem, os grupos são G2, G3, G4, G6, G5, G1 e G7.

2.3 Componentes de variância

Os componentes de variância foram estimados pelo método de máxima verossimilhança residual (PATTERSON; THOMPSON, 1971), e os valores genéticos foram preditos pelo melhor preditor linear não viesado (HENDERSON, 1975), de

acordo com (GILMOUR et al., 2015). Modelos de regressão aleatória foram ajustados através dos polinômios de Legendre, considerando todos os graus possíveis de ajuste para cada efeito aleatório, usando o seguinte modelo:

$$Y_{ijk} = R_k + b_M \phi_{ijM} + \sum_{m=0}^M g_{ikm} \phi_{ijm} + \varepsilon_{ijk}$$

onde Y_{ijk} é o $i^{\text{ésimo}}$ indivíduo ($i = 1, 2, \dots, n$) no $j^{\text{ésimo}}$ grupo ($j = 1, 2, \dots, 7$) na $k^{\text{ésima}}$ repetição ($k = 1, 2, \dots, 10$); R_k é o efeito fixo da repetição; b_M é o coeficiente de regressão fixo ajustado através do sexto grau do polinômio de Legendre para o trajetória média comum dos genótipos. O efeito aleatório, g_{ikm} é o coeficiente de regressão para o polinômio de Legendre de grau m para o valor genético. ϕ_{ijm} é o $m^{\text{ésimo}}$ polinômio de Legendre para a $j^{\text{ésimo}}$ grupo do $i^{\text{ésimo}}$ indivíduo; m é o ajuste do grau, variando de 0 a 6, do polinômio de Legendre para os efeitos genéticos e de ambiente, respectivamente; e ε_{ijk} é o efeito aleatório residual associado a Y_{ijk} .

Na notação de matriz, o modelo acima é descrito da seguinte forma:

$$y = X\tau + Zg + e$$

onde y é o vetor de observações fenotípicas; τ é o vetor dos efeitos de repetição (assumido como fixa); g é o vetor dos efeitos genéticos (assumido como aleatório); e é o vetor de erros (aleatório). X , Z referem-se às matrizes de incidência para esses efeitos.

Neste modelo, $g \sim N(0, K_g \otimes I)$ e $e \sim N(0, R)$; onde K_g é a matriz de covariância para efeitos genéticos; \otimes denota o produto Kronecker; I é uma matriz identidade com ordem apropriada para o respectivo efeito aleatório; e R refere-se à matriz de covariâncias residuais. Diferentes estruturas de covariância residual (homogêneas, diagonais e não estruturadas) foram testadas.

A ordem polinomial em modelos de regressão aleatória foi selecionada usando o critério de akaike (AIC)(SCHWARZ, 1978), como segue:

$$AIC = -2\text{Log}L + 2p$$

onde $\text{Log}L$ é o logaritmo do máximo (L) da função de verossimilhança residual, p é o número de parâmetros estimados.

As estimativas dos componentes de variância (σ_g^2) e os valores genéticos preditos (\tilde{g}_{ij}), na escala original, foram obtidas pelas seguintes expressões (KIRKPATRICK; LOFSVOLD; BULMER, 1990):

$$\sigma_g^2 = \phi_{ijm} k_g \phi_{ijm}$$

$$\tilde{g}_{ij} = \sum_{m=0}^M \alpha_{im} \phi_{ijm}$$

A área abaixo das trajetórias (A), visando ranquear os genótipos, foram obtidas pela seguinte expressão (ALVES et al., 2020):

$$A = \mu + \int_{-1}^1 \sum_{m=0}^M \alpha_{im} \phi_{ijm} x^m dx,$$

Onde μ é a média do fenótipo e x^m é a escala ambiental. As correlações genéticas (ρ_g) entre cada par de grupos ambientais foram obtidos pela seguinte expressão:

$$\rho_g = \frac{\hat{\sigma}_{g(ij)}}{\sqrt{\hat{\sigma}_{g(i)}^2 \hat{\sigma}_{g(j)}^2}}$$

onde $\hat{\sigma}_{g(ij)}$ é a covariância genética entre os genótipos para o par de grupos ambientais i e j ; $\hat{\sigma}_{g(i)}^2$ e $\hat{\sigma}_{g(j)}^2$ são as variâncias genéticas entre genótipo e os grupos ambientais i e j , respectivamente. As análises estatísticas foram realizadas usando os softwares ASReml 4.1 (GILMOUR et al., 2015) e R (R DEVELOPMENT CORE TEAM, 2011).

3. RESULTADOS

Os resultados obtidos no presente estudo favoreceram ao uso dos grupos ambientais para estudo de $G \times E$ na cultura da Soja (Tabela 1). A escolha do polinômio de Legendre de sexta ordem para descrever os efeitos genéticos para a característica produtividade, assumindo a estrutura de resíduo como diagonal, apresentou o melhor resultado (menor valor) pelo critério de Akaike (AIC). Assim, este modelo foi adotado para estimar os componentes de variância e prever os valores genéticos.

A variância genética não foi constante ao longo dos ambientes (Tabela 2). A maior estimativa é do grupo 6 (143695.49) e a menor do do grupo 1 (23119.00). Além

disso, a variância residual no grupo 6 também foi a maior (909701.00), enquanto a menor foi do grupo 2 (386282.00). Para a variância fenotípica, o maior valor foi do grupo 6 (1053396.49), e o menor do grupo 2 (450079.85). Conseqüentemente, a herdabilidade varia de 0.0472 (grupo 1) a 0.2253 (grupo 5).

Tabela 1. Resultados dos modelos convergidos da regressão aleatória para a característica produtividade de grãos avaliados com 20 genótipos de soja.

| Modelo | Resíduo | GL | Ordem | Parâmetros | | | LogL | AIC |
|--------|---------|-----|-------|------------|------|----|---------|----------|
| | | | | Gen. | Res. | p | | |
| 1 | H | 523 | 0 | 1 | 1 | 2 | -3437.7 | 6879.2 |
| 2 | H | 523 | 2 | 6 | 1 | 7 | -3434.3 | 6882.6 |
| 3 | H | 523 | 3 | 10 | 1 | 11 | -3433.8 | 6889.5 |
| 4 | D | 523 | 0 | 1 | 7 | 8 | -3419.6 | 6855.2 |
| 5 | D | 523 | 4 | 15 | 7 | 22 | -3423.1 | 6890.3 |
| 6 | D | 523 | 6 | 28 | 7 | 35 | 6592.2 | -13115.0 |
| 7 | NE | 523 | 0 | 1 | 28 | 29 | -5058.8 | 10175.7 |
| 8 | NE | 523 | 4 | 15 | 28 | 43 | -5055.6 | 10197.1 |
| 9 | NE | 523 | 6 | 28 | 28 | 56 | 4964.9 | -9817.7 |

H: Resíduo Homogêneo; D: Resíduo diagonal; NE: Resíduo não estruturado; Ordem: Ordem do polinômio de Legendre; GL: Graus de liberdade; Gen: Número de parâmetros genéticos; Res: Número de parâmetros residuais; p: Número total de parâmetros; LogL: Logaritmo da função de máxima verossimilhança restrita.

As diferentes normas de reação exibiram diferentes comportamentos de curva, apresentando interação genótipo-ambiente do tipo complexa para o conjunto de 20 genótipos de Soja (Figura 1).

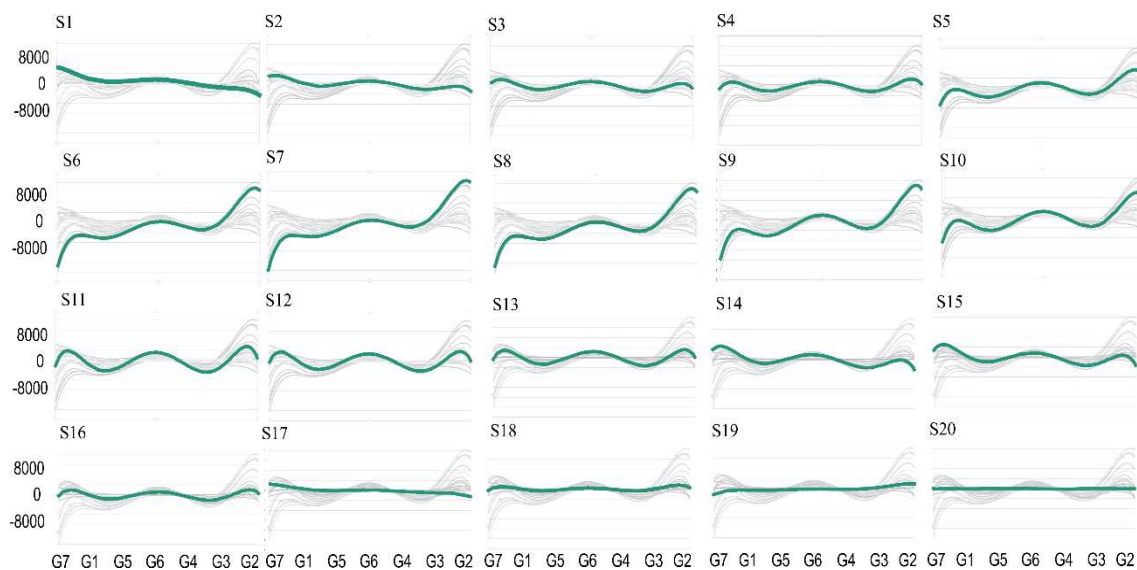


Figura 1. Trajetórias das normas de reação dos 20 genótipos avaliados. S_n são os códigos do genótipos; eixo horizontal com os grupos ambientais apresentados no capítulo 1 (G1,G2...G7); eixo vertical com a escala dos valores genéticos.

Tabela 2. Estimativas para os componentes de variância e parâmetros genéticos.

| Componentes de Variância | Gradiente ambiental | | | | | | |
|--------------------------|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 1 Grupo 7 | 2 Grupo 1 | 3 Grupo 5 | 4 Grupo 6 | 5 Grupo 4 | 6 Grupo 3 | 7 Grupo 2 |
| σ_g^2 | 74500.22 | 23119.00 | 112921.59 | 143695.49 | 76095.16 | 62339.41 | 63797.85 |
| σ_e^2 | 630265 | 466846 | 388257 | 909701 | 648163 | 586922 | 386282 |
| σ_f^2 | 404765.22 | 489965.00 | 501178.59 | 1053396.49 | 724258.16 | 649261.41 | 450079.85 |
| h^2 | 0.1057 | 0.0472 | 0.2253 | 0.1364 | 0.1051 | 0.0960 | 0.1417 |

Grupo: número relacionado ao grupo ambiental do capítulo 1; σ_g^2 variância genética; σ_e^2 variância residual; σ_f^2 variância fenotípica; h^2 herdabilidade.

As áreas sob as trajetórias das normas de reação foram calculadas para o ranqueamento dos genótipos, e os maiores valores representam os genótipos com o melhor desempenho geral ao longo dos ambientes. Neste caso, diferentes valores das áreas foram encontrados, variando de -64539 a 30403.5, e os cinco genótipos com melhor desempenho geral foram: S1, S8, S12, S20 e S9; o genótipo com o menor desempenho geral foi o S6.

As correlações genéticas entre pares de ambientes apresentaram magnitudes moderadas ($0,33 < \rho_g < 0,66$) em G7-G1, G7-G2, G1-G5, G6-G3 e G4-G2; altas magnitudes ($0,67 < \rho_g < 0,89$) em G7-G4, G7-G3, G1-G6, G1-G4, G1-G3, G5-G4, G5-G3, G5-G2, G6-G2 e G4-G3; e magnitudes muito altas ($\rho_g > 0,90$) entre G7-G5 (Figura 2).

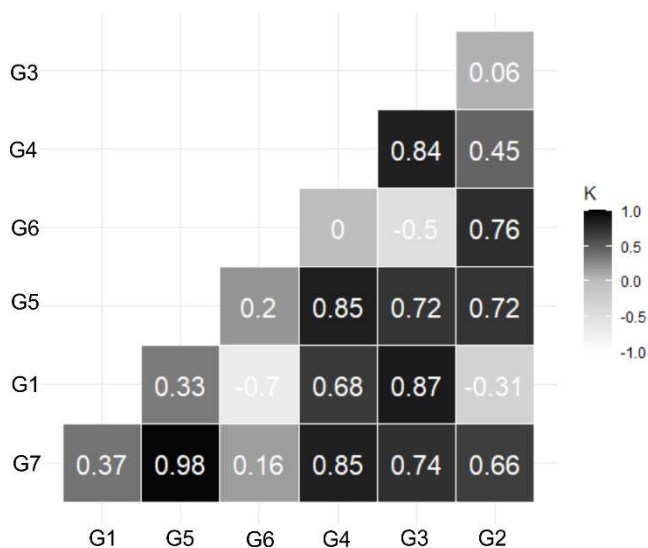


Figura 2. Correlação genética dos valores estimados entre os pares dos grupos ambientais dos 20 genótipos avaliados para produtividade.

4. DISCUSSÃO

No melhoramento genético, a regressão aleatória apresenta grande utilidade, pois permite a predição de valores genéticos de indivíduos avaliados em diferentes anos, locais, ambiente comum, com efeito de ordenação e seleção (SCHAEFFER, 2004). As funções de covariância podem expressar de maneira mais realística, os fenômenos associados a dados longitudinais, sendo superiores a modelos de repetibilidade e multicares (MEYER, 1998). Além disso, polinômios de Legendre

tem sido usado para modelar curvas de crescimento de plantas perenes (LI et al., 2017).

Analisando individualmente cada ambiente do presente estudo, os ambientes G2 e G1 apresentaram correlação genética negativa (Figura 2). Segundo Lande et al. (1983) valores negativos de correlação implicam na ação de diferentes expressões gênicas de um mesmo genótipo. A recomendação e seleção de genótipos para esses locais deve ser feita individualmente, esse fato pode ser observado na figura 1. As melhores adaptabilidades ranqueadas para essas regiões foram as cultivares S1 para o ambiente G7 e S7 para o G2. É importante observar que estes mesmos genótipos tiveram expressões gênicas diferentes quando avaliados nos ambientes G2 e G7.

Para os ambientes G3 e G1, percebe-se que o comportamento das normas apresentam pontos de mínima, denominados de vales. Para estes ambientes, o desempenho genético para as cultivares utilizadas foram menores, assim, recomenda-se a indicação de genótipos mais estáveis como S1, S17 e S20. Quando analisado a interação genótipos por anos, como é feito pela comparação dos grupos G1-G6, as cultivares tiveram comportamentos genéticos distintos, sendo comprovado pela alta correlação negativa (-0,7) entre estes. Para a recomendação eficiente de genótipos, explorando o máximo potencial da adaptabilidade individual, é necessário entender o comportamento climático desta região.

Ao se analisar a interação Genótipos x Locais, como comparado os grupos classificados no Capítulo 1 com mudança apenas na latitude, como G5 e G7 (2018-2019) e os mesmos grupos G1 e G2 (2020), as correlações genéticas foram de 0,2 e -0,31. Estes valores sugerem a necessidade do uso de cultivares com grupos de maturidade relativa distintas.

As trajetórias genéticas apresentadas em forma de normas de reação descrevem os valores genéticos de cada cultivar ao longo do gradiente ambiental. O modelo de regressão aleatória pode prever o valor genético para qualquer cultivar para qualquer grupo ambiental (entre o primeiro e o último grupo). As trajetórias demonstraram que as cultivares tiveram desempenhos semelhantes no grupo 7 e grupo 5, o que revela que a recomendação de cultivos para essas regiões podem ser semelhantes. As correlações genéticas reforçam a eficiência na recomendação. Além disso, o grupo 5 apresentou maiores valores de σ_g^2 e de h_g^2 , sendo assim, um ambiente mais propício na prática de seleção de cultivares, principalmente por

apresentar correlações genéticas de magnitudes moderadas a alta com os grupos ambientais G2, G3 e G4, respectivamente.

As trajetórias genéticas das curvas das 20 cultivares de soja reforçam a presença de interação genótipo x ambiente, uma vez que suas trajetórias são não-lineares e se cruzam, o que implica em uma classificação diferente a cada ambiente. Além disso, as trajetórias também podem ser interpretada como variabilidade genética. Quanto mais distantes as trajetórias genéticas umas das outras, mais geneticamente distintas são os genótipos (GOMULKIEWICZ; KIRKPATRICK, 1992).

Neste estudo, o ranking de genótipos foi realizado com base nas áreas sob as trajetórias genéticas. Portanto, a seleção genética foi baseada na área sob as trajetórias genéticas. A vantagem desta estratégia é que a resposta de seleção pode ser prevista, não apenas na expressão do genótipo sujeito a qualquer ambiente, mas também na quantificação da sensibilidade ambiental através das trajetórias genéticas, ou seja, baseada na capacidade de resposta às mudanças do ambiente (ALVES et al., 2020).

5. CONCLUSÃO

O modelo de regressão aleatória pode ser usado para descrever o comportamento dos genótipos ao longo do gradiente, de forma eficiente, e permite a seleção com base na adaptabilidade e estabilidade dos materiais. Assim, nossos resultados sugerem que modelos de regressão aleatória ajustados através de polinômios de Legendre podem ser usados de forma eficiente em estudos de interação genótipo-ambiente na cultura da soja.

6. REFERÊNCIAS

- ALLIPRANDINI, L. F. et al. Understanding soybean maturity groups in brazil: Environment, cultivar classification, and stability. **Crop Science**, 2009.
- ALVES, R. S. et al. Optimization of Eucalyptus breeding through random regression models allowing for reaction norms in response to environmental gradients. **Tree Genetics & Genomes**, 2020.
- BOURRET, A. et al. Multidimensional environmental influences on timing of breeding

- in a tree swallow population facing climate change. **Evolutionary Applications**, 2015.
- BRADSHAW, J. E. Plant breeding: past, present and future. **Euphytica**, v. 213, n. 3, p. 1–12, 2017.
- BRAWNER, J. T. et al. Classifying genotype by environment interactions for targeted germplasm deployment with a focus on Eucalyptus. **Euphytica**, 2013.
- BRAWNER, J. T. et al. Visualising the environmental preferences of pinus tecunumanii populations. **Tree Genetics and Genomes**, 2014.
- CALUS, M. P. L.; BIJMA, P.; VEERKAMP, R. F. Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model. **Genetics Selection Evolution**, 2004.
- COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, 1994.
- DA SILVA JUNIOR, C. A. et al. Correlations and genetic parameters between morphological descriptors in soybean. **Journal of Agronomy**, 2014.
- DING, M. et al. **Multi-environment trial analysis for Pinus radiata**. New Zealand Journal of Forestry Science. **Anais...**2008
- DUARTE, J. B.; VENCOSKY, R. Estimaco e predico por modelo linear misto com ênfase na ordenaco de mdias de tratamentos genticos. **Scientia Agricola**, 2001.
- EBERHART, S. A.; RUSSELL, W. A. Stability Parameters for Comparing Varieties 1 . **Crop Science**, 1966.
- EL-SODA, M. et al. **Genotype × environment interaction QTL mapping in plants: Lessons from Arabidopsis**Trends in Plant Science, 2014.
- EMBRAPA SOJA. **Tecnologias de Produo de Soja - Regio Central do Brasil 2020**. [s.l: s.n.].
- GIANOLA, D. et al. Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. **BMC Genetics**, 2011.
- GILMOUR, A R. et al. ASReml User Guide Release 4.1 Structural Specification. **VSN International Ltd. Hemel Hempstead**, 2015.
- GOMULKIEWICZ, R.; KIRKPATRICK, M. Quantitative genetics and the evolution of reaction norms. **Evolution**, 1992.
- GRAY, L. K. et al. Developing management strategies for tree improvement programs under climate change: Insights gained from long-term field trials with lodgepole pine.

Forest Ecology and Management, 2016.

HAYKIN, S. Redes neurais: princípios e prática. **Bookman**, 2017.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, 1975.

KASTER, M.; FARIAS, J. R. B. **REGIONALIZAÇÃO DOS TESTES DE VALOR DE CULTIVO E USO E DA INDICAÇÃO DE CULTIVARES DE SOJA**. XXIV Reunião de Pesquisa de Soja da Região Central do Brasil. **Anais...**2002

KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics**, 1990.

KOHONEN, T. The Self-Organizing Map. **Proceedings of the IEEE**, 1990.

KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, 2013.

KOLMODIN, R.; BIJMA, P. Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. **Genetics Selection Evolution**, 2004.

LANDE, R.; ARNOLD, S. J. The Measurement of Selection on Correlated Characters. **Evolution**, 1983.

LI, Y. et al. **Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application** **Tree Genetics and Genomes**, 2017.

LYNCH, M., WALSH, B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA, USA. **CEUR Workshop Proceedings**, 2017.

METTLER, L. E.; GREGG, T. G. Population genetics and evolution. In: São Paulo, Polígono, Ed. da USP: [s.n.]. p. 262 p.

MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. **Genetics Selection Evolution**, 1998.

OGUT, F. et al. Genetic analysis of Diallel progeny test data using factor analytic linear mixed models. **Forest Science**, 2014.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, 1971.

R DEVELOPMENT CORE TEAM, R. **R: A Language and Environment for Statistical Computing**. [s.l: s.n.].

RAYMOND, C. A. Genotype by environment interactions for *Pinus radiata* in New South Wales, Australia. **Tree Genetics and Genomes**, 2011.

- RESENDE, R. T. et al. Intra-genotypic competition of Eucalyptus clones generated by environmental heterogeneity can optimize productivity in forest stands. **Forest Ecology and Management**, 2016.
- REZENDE, P. M. DE; CARVALHO, E. DE A. Avaliação de cultivares de soja [Glycine max (L.) Merrill] para o sul de Minas Gerais. **Ciência e Agrotecnologia**, v. 31, n. 6, p. 1616–1623, 2007.
- RITTER, H.; KOHONEN, T. Self-organizing semantic maps. **Biological Cybernetics**, 1989.
- SCHAEFFER, L. R. Application of random regression models in animal breeding. **Livestock Production Science**, 2004.
- SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, 1978.
- SOARES, A. A. V. et al. Increasing stand structural heterogeneity reduces productivity in Brazilian Eucalyptus monoclonal stands. **Forest Ecology and Management**, 2016.
- SOUSA, L. B. et al. Evaluation of soybean lines and environmental stratification using the AMMI, GGE biplot, and factor analysis methods. **Genetics and Molecular Research**, 2015.
- SQUILASSI, M. G. **Interação de genótipos com ambientes**. [s.l: s.n.].
- VESANTO, J., HIMBERG, J., ALHONIEMI, E., AND PARHANKANGAS, J. SOM Toolbox for Matlab 5, Report A57. **Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland**, 2000.
- WEHRENS, R.; BUYDENS, L. M. C. Self- and super-organizing maps in R: The kohonen package. **Journal of Statistical Software**, 2007.
- WOLTERECK, R. Weitere experimentelle untersuchungen über Artänderung, speziell über das Wesen quantitativer Artunterschiede bei Daphniden. **Zeitschrift für Induktive Abstammungs- und Vererbungslehre**, 1913.
- XAVIER, A. et al. Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. **G3: Genes, Genomes, Genetics**, 2018.