

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos quantitativos e classificatórios para avaliar a produtividade de soja

Henry Sako

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Fitotecnia

Piracicaba
2022

Henry Sako
Engenheiro Agrônomo

Modelos quantitativos e classificatórios para avaliar a produtividade de soja
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **JOSÉ LAÉRCIO FAVARIN**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Fitotecnia

Piracicaba
2022

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Sako, Henry

Modelos quantitativos e classificatórios para avaliar a produtividade de soja/ Henry Sako. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2022.

51 p.

Dissertação (Mestrado) - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Análise discriminante 2. Alta produtividade 3. Soja 4. Aprendizado de máquina 5. Fitotecnia I. Título

A todos produtores que desejam mudança em seu sistema de produção, dedico esse trabalho.

AGRADECIMENTOS

Ao Prof. José Laércio Favarin pela confiança e orientação não só nesse trabalho, mas desde os primeiros trabalhos que fiz, no início do Grupo Paces.

Ao Prof. Hasime Tokeshi, o qual deixou “semeado a idéia” para esse trabalho. A motivação desse trabalho surgiu através das inúmeras conversas que tivemos sobre agronomia desde 2002. Lavouras excepcionais sempre merecem ser estudadas.

Ao CESB (Comitê Estratégico Soja Brasil), em especial ao Luiz Antônio da Silva e ao José Erasmo Soares, que apoiaram esse trabalho.

Ao Joao Dantas, Rafael Battisti e Marcelo Alves, parceiros que abraçaram o desafio de entender e replicar a alta produtividade desde 2014.

Ao Armando Parducci, ao Thiago Camargo e aos colegas do IBRA (Instituto Brasileiro de Análises), sempre apoiando a evolução do conhecimento agrônômico.

A todos os produtores, consultores, técnicos que colaboraram com esse importante trabalho

Aos meus pais, Roberto S. Sako e Alice S Sako. Independente do desafio que tinham em cada época, sempre trouxe a mim e meus irmãos o senso da ética, a importância da educação e ter um objetivo de vida. Pilares as quais me incentivou a voltar a universidade após um sólido conhecimento de campo e me desafiar a expandir as fronteiras do conhecimento agrônômico.

À Juliana e à Lídia que me apoiaram nesse trabalho e entenderam minha ausência na vida delas. Foram percorridos 15.181 km acompanhando lavouras de alta produtividade.

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	10
1. INTRODUÇÃO.....	11
2. REVISÃO DE LITERATURA	13
2.1. Complexidade da agricultura.....	13
2.2. Aprendizado de máquina e estatística multivariada	14
3. MATERIAL E MÉTODOS.....	19
4. RESULTADOS E DISCUSSÃO	21
4.1. Análise descritiva do banco de dados.....	21
4.2. Fatores climáticos e da fertilidade de solo na produtividade da soja	21
4.3. Erro Absoluto dos modelos quantitativos de predição	35
4.4. Modelos de classificação	35
5. CONCLUSÕES	45
REFERÊNCIAS	47

RESUMO

Modelos quantitativos e classificatórios para avaliar a produtividade da soja

O aumento de produtividade de uma lavoura é um dos pilares para preservar a sustentabilidade do negócio agrícola. O Desafio Nacional de Máxima Produtividade promovido pelo CESB (Comitê Estratégico Soja Brasil) tem revelado diversas lavouras de produtividades superiores a 5.400 kg.ha⁻¹, bem acima da produtividade média brasileira, que é de 3.000 a 3.600 kg.ha⁻¹. Dessa forma, é uma oportunidade entender os fatores agronômicos envolvidos na alta produtividade de soja, assim como elaborar algoritmos capazes de prever a produtividade de soja com base em variáveis agronômicas, para auxiliar os técnicos a manejar os fatores que levam a soja a expressar seu potencial produtivo. A produtividade de uma lavoura é resultante de um conjunto de fatores agronômicos de clima, solo e manejo e, por esse motivo, é importante o conhecimento das variáveis agronômicas de forma abrangente, para a construção do algoritmo. Foram reunidos dados de 91 lavouras com variáveis de clima, manejo, fertilidade de superfície e subsuperfície do solo e resistência do solo ao crescimento radicular, totalizando 219 variáveis agronômicas. Foram aplicados modelos quantitativos e classificatórios para estimar a produtividade de soja por meio das metodologias Floresta Randômica, Rede Neural, Suport Vector Machine e Análise Discriminante e identificados os fatores agronômicos em comum na alta produtividade de soja. A Análise Discriminante teve um índice de acerto de 91,5%, classificando produtividades acima e abaixo de 5.500 kg.ha⁻¹. Das 219 variáveis agronômicas, o modelo discriminante destacou 7 fatores agronômicos: resistência do solo a 20 cm e 50 cm de profundidade; pH CaCl₂ na profundidade de 10-20 cm e 80-100 cm, saturação de alumínio a 0-10 cm, Fósforo a 0-10 cm; Enxofre a 80-100 cm de profundidade; cobre a 10-20 cm de profundidade; temperatura média do estágio fenológico de R1 a R5, temperatura mínima do estabelecimento à emergência e o déficit hídrico durante o ciclo da soja. O modelo discriminante demonstrou a relevância dos atributos físicos do solo e da fertilidade química na superfície e, principalmente, em subsuperfície para a obtenção de alta produtividade de soja. As metodologias Floresta Randômica, Rede Neural e Support Vector Machine não construíram bons modelos por causa da necessidade de um número maior de dados para aplicação desses modelos.

Palavras-chave: Análise discriminante, Alta produtividade, Soja, Aprendizado de máquina, Fitotecnia

ABSTRACT

Quantitative and qualitative models to evaluate soybean yield

Increasing the productivity of a crop is one of the pillars to preserve the sustainability of the agricultural business. The National Maximum Productivity Challenge promoted by CESB (Brazilian Soybean Strategic Committee) has revealed several crops with yields above 5,400 kg.ha⁻¹, much superior than Brazilian's productivity average, of 3,000 to 3,600 kg.ha⁻¹. Thus, understanding the agronomic factors involved in high soybean productivity it is an opportunity, as well as developing algorithms capable of predicting soybean productivity, based on agronomic variables, to help technicians to manage the factors that lead soybean to express its productive potential. The productivity of a crop is the result of a set of agronomic factors, such as climate, soil and management. Therefore, it is important to know the agronomic variables comprehensively, in order to build the algorithm. Data from 91 crops were collected, with climate, management, soil surface and subsurface fertility and soil resistance to root growth variables, totaling 219 agronomic variables. Quantitative and classificatory models were applied to estimate soybean productivity using the Random Forest, Neural Network, Support Vector Machine and Discriminant Analysis methodologies. The Discriminant Analysis had a 91,5% success rate, classifying productivity above and below 5,500 kg.ha⁻¹. Considering the 219 agronomic variables collected, the discriminating model highlighted 7 agronomic factors: soil resistance at 20 cm and 50 cm depth; pH CaCl₂ at 10-20 cm and 60-80 cm depth; aluminum saturation at 0-10 cm; phosphorus at 0-10 cm; sulfur at 80-100 cm depth; copper at 10-20 cm; medium temperature of soybean phenological stage R1 to R5; minimum temperature from seedling to emergence and water deficit during the soybean cycle. The discriminant model demonstrates the relevance of understanding the physical attributes of the soil and fertility on the surface and, mainly, on the subsurface, to high soybean productivity. The Random Forest, Neural Network and Support Vector Machine methodologies did not build good models considering the data volume withing the collected database.

Keywords: Discriminant analysis, High yield, Soybean, Machine learning, Phytotechnics

LISTA DE FIGURAS

Figura 1. Visão geral do método Floresta Randômica (ARAÚJO, 2013).	15
Figura 2. Representação do neurônio de McCulloch e Pitts (FURTADO, 2019).	15
Figura 3. Representação das camadas de uma rede neural artificial (DA SILVA; SPATTI; FLAUZINO, 2010).	16
Figura 4. Valores de resistência do solo a 20 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha-1) no eixo y.....	23
Figura 5. Valores de resistência do solo a 30 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha-1) no eixo y.....	24
Figura 6. Valores de resistência do solo a 40 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha-1) no eixo y.....	24
Figura 7. Valores de resistência do solo a 50 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha-1) no eixo y.....	25
Figura 8. Teores de fósforo (mg.dm-3, Resina) na camada 0-10 cm no eixo x e produtividade da soja (kg.ha-1) no eixo y.....	26
Figura 9. Teores de fósforo (mg.dm-3, Resina) na camada 10-20 cm no eixo x e produtividade da soja (kg.ha-1) no eixo y.....	27
Figura 10. Relação entre o pH CaCl2 na camada de 0-10 cm de profundidade e a produtividade de soja (kg.ha-1).	29
Figura 11. Relação entre o pH CaCl2 na camada de 10-20 cm de profundidade e a produtividade de soja (kg.ha-1).	30
Figura 12. Relação entre o pH CaCl2 na camada de 20-40 cm de profundidade e a produtividade da soja (kg.ha-1).	30
Figura 13. Relação entre o pH CaCl2 a 60-80 cm de profundidade e a produtividade da soja (kg.ha-1).....	31
Figura 14. Relação entre o pH CaCl2 a 80-100 cm de profundidade e a produtividade da soja (kg.ha-1).....	31
Figura 15. Variáveis canônicas 1 e 2 e a predição da produtividade em ACIMA E ABAIXO Da mediana como divisor do banco de dados.....	37
Figura 16. Variáveis canônicas 1 e 2 e a predição da produtividade em ACIMA E ABAIXO Da, terceiro quartil como divisor do banco de dados, 5.500 kg.ha-1.....	39

Figura 17. A) infiltração de uma solução com esmalte sintético branco numa lavoura de produtividade de 50 sc.ha-1 (SAKO, 2019). B) infiltração de água numa lavoura com produtividade maior que 100 sc.ha-1 (CESB, 2021).....	41
Figura 18. Teores de enxofre (mg.dm-3) na camada de 80-100 cm de profundidade e produtividade da soja (kg.ha-1).....	42
Figura 19. Teores de cobre (mg.dm-3, DTPA) na camada de 10-20 cm de profundidade em relação à produtividade (kg.ha-1).....	43
Figura 20. Déficit hídrico (mm de água) da semeadura até R8 e a penalização na produtividade da soja (kg.ha-1).....	44

LISTA DE TABELAS

Tabela 1. Estatística descritiva do banco de dados reunido na dissertação.	21
Tabela 2. Aspectos sobre as variáveis agronômicas usadas, número de cada uma delas e o grau de influência na produtividade (%).	21
Tabela 3. Correlação de Pearson das variáveis com maior influência na produtividade. Legenda da coluna F: (M) variáveis de solo alteráveis pelo manejo a 20 cm; (MP) variáveis manejáveis abaixo de 20 cm de profundidade; (S) variáveis difíceis de serem alteradas por manejo ou não possíveis de serem alteradas; (C) fatores associados ao clima. De todas as variáveis agronômicas, as 20% de maior relevância em relação à produtividade estão destacadas em negrito. As siglas acompanhadas nas variáveis climáticas representam o período do estágio fenológico.	22
Tabela 4. Valores de resistência do solo (Res) em dois grupos de produtividades de soja: abaixo e acima de 4.920 kg.ha-1.	25
Tabela 5. Teores de fósforo (P, mg.dm-3), potássio (K, mmolc.dm-3), magnésio (Mg, mmolc.dm-3), cálcio (Ca, mmolc.dm-3) em produtividades acima e abaixo de 4.920 kg.ha-1, teste t.	28
Tabela 6. Valores de pH CaCl2 em produtividades acima e abaixo de 4.920 kg.ha-1.	29
Tabela 7. Porcentagem de potássio, magnésio e cálcio na CTC em produtividades acima e abaixo de 4.920 kg.ha-1.	32
Tabela 8. Fertilidade do perfil de solo entre produtividades acima e abaixo de 4.920 kg.ha-1, os valores foram delimitados pelo teste t a 5% de significância para saturação por bases (V%) enxofre (S), matéria orgânica (MO) e saturação de alumínio (m%).	33
Tabela 9. Valores de nutrientes estabelecidos pelo teste t em produtividades abaixo e acima de 4.920 kg.ha-1. Boro (B); Cobre (Cu) Zinco (Zn) Manganês (Mn) Ferro (Fe).	34
Tabela 10. Erro absoluto médio do modelo avaliado.	35
Tabela 11. Porcentagem de acerto em classificação tendo o valor da mediana, 4.920 kg.ha-1 como classificador dos valores.	35
Tabela 12. Variáveis selecionadas no modelo da Análise Discriminante, tendo a mediana como divisor do banco de dados, no valor de 4.950 kg.ha-1.	36
Tabela 13. Relação canônica com as variáveis agronômicas.	38

1. INTRODUÇÃO

Desde 2008, o Desafio Nacional de Máxima Produtividade, promovido pelo CESB, tem revelado nacionalmente diversas lavouras de produtividade acima de $5.400 \text{ kg}\cdot\text{ha}^{-1}$ (CESB, 2017) enquanto, nesse período, a produtividade média brasileira tem oscilado entre 3.000 a $3.600 \text{ kg}\cdot\text{ha}^{-1}$, instigando os técnicos a entenderem as variáveis agronômicas que têm limitado a produtividade. Somado a isso, o lucro dos produtores tem sido pressionado pela inflação dos custos de produção e pela redução dos preços das commodities agrícolas, processo conhecido como o “declínio secular” dos preços (LEITE; WAQUIL, 2006; MAZOYER; ROUDART, 1997). Apesar do tema produtividade ser um motivo de estudo por séculos entre os técnicos, ainda há um longo trabalho de pesquisa a ser feito em torno desse assunto quando se trata de entender o conjunto de fatores agronômicos responsáveis pela alta produtividade ou estimar a produtividade de uma lavoura por meio dos fatores agronômicos. A complexidade do tema é evidente mesmo quando se analisa a diferença entre talhões adjacentes de uma lavoura de alta e baixa produtividade. Malavolta (2006) cita que há 56 fatores que influenciam na produtividade de soja. Em 1982, Cooke já alertava sobre a importância de entender a interação das variáveis agronômicas para sair da estagnação da produtividade (COOKE, 1982). Ruffo et al. (2010) avaliaram 5 fatores para influenciar a produtividade de milho e puderam verificar que a mudança de híbrido aumentou a produtividade em $460 \text{ kg}\cdot\text{ha}^{-1}$; a adubação nitrogenada em $290 \text{ kg}\cdot\text{ha}^{-1}$; a adubação fosfatada mais micronutriente aumentou em $330 \text{ kg}\cdot\text{ha}^{-1}$; o uso de defensivos em $450 \text{ kg}\cdot\text{ha}^{-1}$ e a população reduziu $240 \text{ kg}\cdot\text{ha}^{-1}$. A soma isolada delas levaria a um aumento de $1.290 \text{ kg}\cdot\text{ha}^{-1}$, no entanto, quando avaliaram-se todos os fatores em conjunto, aumentou em $2.870 \text{ kg}\cdot\text{ha}^{-1}$, tornando-se evidente que a interação dos fatores agronômicos é mais importante que a influência desses fatores de forma isolada.

Os métodos estatísticos “tradicionais” possuem limitações quando processam problemas compostos por diversas variáveis de natureza não-linear, como muitas vezes é a produtividade. Nesse sentido, os métodos de análise multivariada têm se apresentado como excelente proposta para entender problemas dessa natureza. Hammer (2016) comparou 4 diferentes modelos para estimar a produtividade de cana-de-açúcar: Floresta Randômica, Boosting, Support Vector Machine e o modelo agrometeorológico convencional. Nos três primeiros modelos, obteve um erro do quadrado médio (RMSE) de 19.700 a $20.030 \text{ kg}\cdot\text{ha}^{-1}$, um desempenho $34.000 \text{ kg}\cdot\text{ha}^{-1}$ superior ao modelo agrometeorológico. Alves (2016) utilizou a Rede Neural Artificial para estimar a produtividade de soja por meio de parâmetros biométricos e obteve um índice de acerto de 72%. Kaul e colaboradores (2005) aplicaram a metodologia da Rede Neural para prever a produtividade na região de Maine, Estados Unidos, e com dados de clima e solo encontrou uma correlação de 77% entre a produtividade estimada e a obtida. Smidt e colaboradores (2016), utilizando a metodologia Floresta Randômica com dados de macronutrientes do solo, capacidade de armazenamento de água e clima, conseguiram organizar os fatores que mais limitam a produtividade de soja em 22 locais dos Estados Unidos. Dessa forma, as metodologias de análise de dados multivariados e de aprendizado de máquina se mostram como uma ferramenta poderosa para construir um modelo para prever a produtividade de soja.

O Aprendizado de Máquina (Machine Learning) possui diferentes métodos de análise. A Floresta Randômica é a junção de um conjunto de árvores de decisões em que cada variável analisada separa-se em duas possibilidades e repete-se sucessivamente até se chegar à predição de um determinado resultado (GERON, 2017). Já as Redes Neurais são compostas por camadas, de forma que a entrada de duas ou mais variáveis do problema avaliado se integram em uma ou mais camadas, multiplicadas por um coeficiente calculado pela Rede Neural e que posteriormente são agrupadas e é calculado um valor de predição (GERON, 2017).

Uma outra análise estatística multivariada que tem grande potencial de uso na agricultura é a análise discriminante. Essa análise consiste em classificar em dois ou mais grupos o problema avaliado por meio de dados qualitativos ou quantitativos transformados em qualitativos, tomando como critério diversas variáveis e com a vantagem de ser uma análise supervisionada e clara. Uma vantagem dessa análise é uma menor exigência de banco de dados e é uma metodologia útil quando se tem o problema de dimensionalidade, ou seja muitas variáveis envolvidas na análise (LIAKOS et al., 2018).

O aumento da produtividade é um tema relevante para a sustentabilidade das lavouras, pois, como citado anteriormente, a lucratividade ficaria prejudicada em caso contrário, em virtude do declínio secular dos preços. Por outro lado, são pouco frequentes os registros de alta produtividade e as pesquisas em torno dela, o que torna relevante a dedicação científica em torno da alta produtividade de soja. A produtividade possui uma natureza multidisciplinar e essa dissertação visa à construção de algoritmos capazes de estimar a alta produtividade de soja por meio de metodologias multivariadas, a partir de um levantamento abrangente de variáveis agronômicas como clima, solo e manejo, em diversos patamares de produtividade, e a identificação dos fatores agronômicos associados à alta produtividade.

Assim, o objetivo do presente trabalho foi a avaliação de modelos multivariados para estimar a alta produtividade de soja e os fatores agronômicos associados.

2. REVISÃO DE LITERATURA

2.1. Complexidade da agricultura

Na safra de 2016/17 o Brasil produziu 103 milhões de toneladas de soja em 33 milhões de hectares (CONAB, 2017), sendo uma das principais commodities do país. Nos últimos 10 anos, a produtividade nacional tem oscilado entre 2.816 a 3.364 kg.ha⁻¹ de soja (CONAB, 2017) enquanto seu potencial produtivo está muito além desses valores obtidos (BATTISTI, 2016). O Desafio Nacional de Máxima Produtividade, promovido pelo CESB, teve na etapa 2016/2017 um quantitativo de 6.000 lavouras inscritas e, 10% dessas (600 lavouras), tiveram produtividade acima de 5.400 sc.ha⁻¹, ou seja, são escassos os registros de lavouras de alta produtividade.

Diversos fatores compõem a produtividade, desde aspectos do clima, solo e genética até o manejo. A produtividade potencial de uma cultura é influenciada por fatores como radiação solar, temperatura, amplitude do dia, cultivar e ciclo da cultura. Quando se integram aos modelos as variáveis relacionadas à água, como capacidade de água disponível do solo e a disponibilidade de água (precipitação menos a evapotranspiração) e a fertilidade do perfil do solo, forma-se a produtividade atingível. A produtividade real é aquela que considerada todos os fatores citados somados ao manejo da lavoura (SENTELHAS et al., 2017). Sentelhas e colaboradores (2016), avaliando as perdas de produtividade em diversos locais do país, constataram que a perda oriunda por estresse hídrico pode ser de 51 a 81% e a perda por manejo é de 19 a 41%. Uma vez que o estresse hídrico envolve componentes de clima (temperatura, radiação, distribuição e volume de precipitação), fertilidade do perfil do solo (devido à sua influência no crescimento radicular e água disponível), tornam-se variáveis importantes a serem consideradas para entender a produtividade da soja.

A produtividade é influenciada por diversos fatores, entre eles a nutrição mineral de plantas. A nutrição e adubação compõem uma parte importante da produtividade agrícola e, por esse motivo, também estão relacionadas com a produtividade de soja. Para se produzir 6.000 kg.ha⁻¹ de soja são necessários 492 kg.ha⁻¹ de nitrogênio, 43,8 kg.ha⁻¹ de fósforo, 204 kg.ha⁻¹ de potássio, 96 kg.ha⁻¹ de cálcio, 60 kg.ha⁻¹ de magnésio e 27 kg.ha⁻¹ de enxofre (OLIVEIRA et al., 2014). Por diversas razões, para se alcançar altas produtividades, é necessário, ao plantar, acessar esse volume de nutrientes, e a fertilidade do solo é o início de todo o processo.

O levantamento da fertilidade de solo nos horizontes superficiais e subsuperficiais, nos aspectos químicos-físicos-biológicos do solo, são importantes para análise da produtividade das lavouras. Santi (2007) avaliou a fertilidade do solo até 20 cm quanto à disponibilidade de elementos e outros aspectos do solo como a qualidade física. As variáveis que mais explicaram a produtividade foi a qualidade física do solo e os teores de potássio de 0-20cm de profundidade. Sako et al. (2016) realizaram o levantamento de solos até 100 cm de profundidade e outras variáveis agronômicas, e concluíram que a fertilidade de potássio até 100 cm de profundidade, cálcio e magnésio até 40cm foram variáveis importantes que explicaram a produtividade, ou seja, a fertilidade de camadas mais profundas do solo esteve associada à produtividade. Kaul (2005) e colaboradores, aplicando a Rede Neural para fazer a predição da produtividade de soja, obtiveram uma correlação de 77% entre a estimada e a obtida e uma das variáveis analisadas foi um índice de fertilidade do solo que considera atributos do solo até 1 m de profundidade. Em ensaios conduzidos em Iracemápolis, SP, Sako et al. (2017) fizeram a correção de solo a 50, 100, 150 e 200cm e obtiveram incrementos significativos de produtividade à medida que foi aumentando a profundidade de correção, ressaltando dessa forma, a importância da fertilidade em subsuperfície de grandes profundidades para a produtividade de soja. Para estimar a produtividade da lavoura, além da abrangência do levantamento da fertilidade do solo, é importante

reunir outras variáveis agronômicas da forma mais ampla possível, permitindo assim construir bons modelos para a estimativa da produtividade.

A produtividade possui uma natureza multidisciplinar e as variáveis de grande relevância devem ser compostas para construir um modelo de predição de produtividade. Essa dissertação trabalhou com dados de clima, solo e manejo, com uma amplitude grande de produtividade, para construir os fatores em comum da alta produtividade e algoritmos de classificação da alta produtividade.

2.2. Aprendizado de máquina e estatística multivariada

Os modelos de aprendizado de máquina têm ampla possibilidade de aplicações na agricultura, podendo incluir desde a identificação de plantas daninhas, doenças e pragas e classificação de solos, à predição de produtividade. O Aprendizado de Máquina é um conjunto de metodologias que formam um programa que permite que o mesmo consiga aprender com os dados. Mitchel (1997) fez a seguinte definição do aprendizado de máquina: *“O aprendizado de máquina é quando há um aprendizado de uma experiência E com respeito a um tópico T e medido por um método P , sua performance no T por meio de avaliações por P , melhoram com a experiência E ”*. Nessa dissertação foram escolhidas Floresta Randômica, Rede Neural, Suport Vector Machine e Análise Discriminante para construção dos modelos quantitativos-qualitativos para prever a produtividade da soja.

A construção de uma árvore de decisão baseia-se na divisão de um problema analisado em duas possibilidades, de acordo com uma variável ou atributo, e essas divisões serão novamente divididas em outras possibilidades de acordo com o atributo ou variável analisada. Esse processo se repete até que todos os atributos do banco de dados sejam analisados e geradas as predições de resultados. Cada atributo usado para dividir ou classificar o problema é denominado de nó. Dessa forma, um problema maior é decomposto em diversas divisões até se chegar à predição do resultado final, compondo dessa forma a árvore de decisão.

A vantagem das árvores de decisão é a construção de modelos de fácil interpretação, podendo assim serem analisados por um conjunto de regras, além da capacidade de avaliar se as regras geradas são admissíveis (MITCHELL, 1997). Apesar da facilidade de interpretação dos resultados de uma árvore de decisão, ela possui um risco de ter overfitting, ou seja, se tornar muito específica para determinado banco de dados e reduzir sua capacidade de fazer predições, quando a árvore de decisão possui muitos nós, ou é muito grande. Para reduzir esse problema, a junção de diversas árvores de decisão permite gerar duas ou mais predições e, por um sistema de votação, selecionar os melhores resultados, formando o conceito de Floresta Randômica (GERON, 2017). Na Figura 1, apresenta-se um modelo de Floresta Randômica. Segundo Breiman (2001), Floresta Randômica é um classificador composto por uma coleção de árvores $T_k(x)$, $k = 1, 2, \dots, L$ onde T_k são amostras aleatórias independentes e identicamente distribuídas e cada árvore vota na classe mais popular para a entrada x .

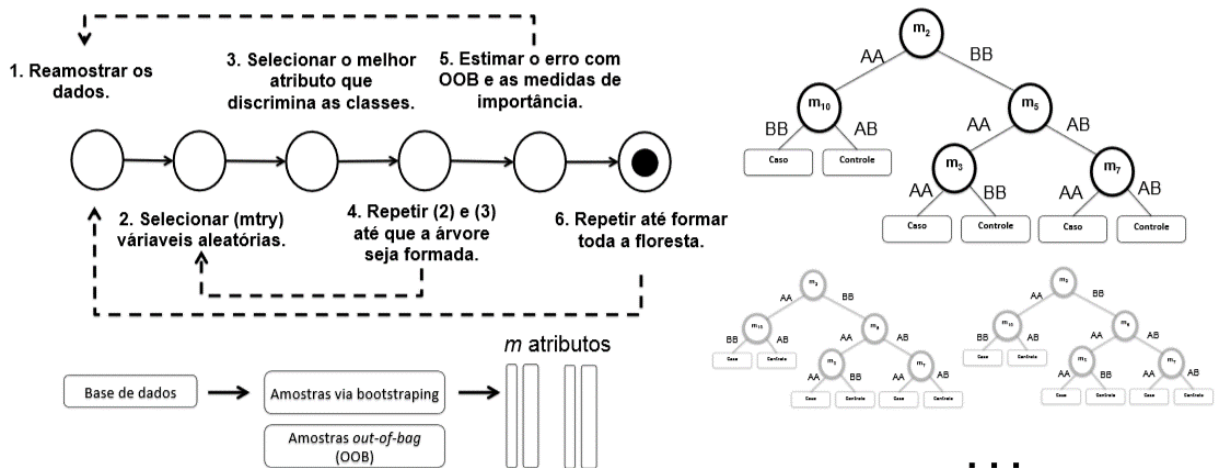


Figura 1. Visão geral do método Floresta Randômica (ARAÚJO, 2013).

Outra metodologia que tem sido amplamente aplicada nos mais diversos campos da ciência é a de Redes Neurais. Em grande parte, isso se deve à sua capacidade de aprendizagem e generalização, permitindo a modelagem de sistemas altamente não-lineares, em que a relação entre as variáveis é desconhecida ou muito complexa (ALEKSANDER; MORTON, 1995; RUSSELL; PETER, 2010).

As Redes Neurais Artificiais (RNA), também conhecidas como connexionismo ou sistemas de processamento paralelo e distribuído, são estruturas baseadas em modelos que lembram a estrutura do cérebro humano, que utilizam unidades de processamento simples (neurônios artificiais) que calculam determinadas funções matemáticas para criar conexões que armazenam o conhecimento das RNAs (FURTADO, 2019).

A descrição matemática de neurônio artificial, proposto por McCulloch e Pitts (1943), segue um modelo composto por n entradas (terminais, equivalente aos dendritos) que são responsáveis por receber os valores x_1, x_2, \dots, x_n e apenas um terminal de saída y (equivalente ao axônio). O comportamento das sinapses é simulado pela aplicação de pesos (w_1, w_2, \dots, w_n), onde as entradas podem assumir valores positivos ou negativos, dependendo das sinapses correspondentes, serem inibitórias ou excitatórias. A saída só é acionada quando a soma dos impulsos recebidos (Σ) ultrapassa o limiar de excitação $[g(u)]$ (Figura 2).

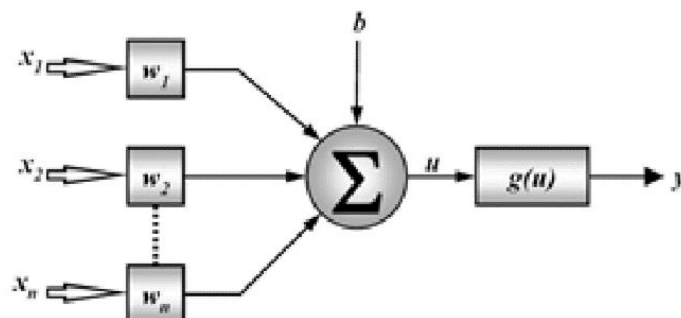


Figura 2. Representação do neurônio de McCulloch e Pitts (FURTADO, 2019).

O modelo matemático que descreve o comportamento do neurônio artificial é expresso pela equação:

$$y = g \left(\sum_{i=1}^n w_i \cdot x_i + b \right)$$

em que: n corresponde ao número de entradas do neurônio; w_i é o peso associado com a i -ésima entrada; b é o limiar associado ao neurônio; x_i é a i -ésima entrada do neurônio; y é a saída do neurônio.

Descrita a equação, um neurônio artificial opera da seguinte forma: a) os sinais são apresentados à entrada; b) cada sinal é multiplicado por um peso que representa o seu nível de relevância na saída da unidade; c) é feita uma soma ponderada dos sinais, resultando assim em um nível de atividade; d) caso este nível de atividade exceda certo limite (threshold), então o neurônio produz uma saída.

O conceito da Rede Neural é quando se reúnem diversos neuronios artificiais. Podemos ilustrar a estrutura de uma rede neural, basicamente, em três partes (Figura 3): camada de entrada - responsável pelo recebimento das informações de entrada, isto é, onde os padrões são inicialmente apresentados à rede; camadas intermediárias ou ocultas - responsáveis pela extração das características do processo ou sistema ao qual se está inferindo, é a camada onde praticamente todo o processamento da rede é realizado; camada de saída - responsável pela apresentação do resultado alcançado, o qual foi processado através de todas as camadas da rede.

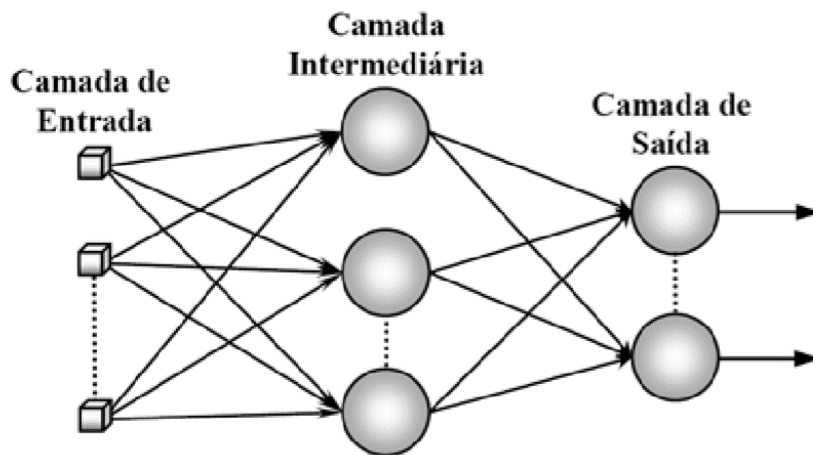


Figura 3. Representação das camadas de uma rede neural artificial (DA SILVA; SPATTI; FLAUZINO, 2010).

Uma outra análise que tem sido aplicada na agricultura é a análise discriminante. A análise discriminante foi proposta por Fisher em 1936 como um método para classificar uma população em uma ou duas categorias (TIMM, 2002). A análise discriminante constrói duas ou mais equações para cada categoria a ser classificada, contemplando as variáveis propostas, a um determinado nível de significância. Perante a soma dos resultados das duas equações, se chega a duas possibilidades:

Se: Soma > 0 = classificação A; se a Soma < 0 = classificação B

O uso da análise discriminante na agricultura tem abrangido diversas aplicações como classificação de solos, identificação de plantas daninhas, doenças de plantas e produtividade (LIAKOS et al., 2018). Voloudakis et al. (2015) aplicou a análise discriminante na produção do algodão para identificar se havia perda de produtividade por déficit hídrico e manejar a irrigação. Cox & Gerard (2010) fizeram aplicação da análise discriminante para identificar zonas de manejo a partir de dados de textura, pH, nutrientes e teores totais de carbono e nitrogênio. Wu et al. (2004) aplicaram a análise discriminante para avaliar os riscos de perda de produtividade em soja e milho por déficit hídrico, em soja o índice de assertividade foi de 80% quando a soja começou o estágio fenológico de enchimento de grãos. Jaynes et al. (2003), adotando a análise discriminante para identificar clusters de diferentes produtividades de milho, analisaram dados de condutividade elétrica, altitude, declividade e conseguiram 76 a 80% de taxa de sucesso na classificação da produtividade.

3. MATERIAL E MÉTODOS

De maio de 2014 a junho de 2021 foi levantado um conjunto de dados pelo Comitê Estratégico Soja Brasil (CESB) e pela empresa DK Ciência Agronômica, somando um volume de número de lavouras com dados de clima, solo, manejo e produtividade de 91 locais. Os locais de coleta estão distribuídos entre os estados de MG, GO, MT, MS, PR, RS e SP e todos dados foram levantados em Latossolos Vermelho para evitar grandes variações do tipo de solo e foram avaliados cobrindo uma unidade de avaliação de 2,5 a 3 hectares. Muitas dessas lavouras receberam investimentos para expressar a produtividade e participaram do Desafio Nacional de Máxima Produtividade promovido pelo CESB. Os dados levantados foram:

i. Dados de clima: Foram realizados monitoramento diário de precipitação durante o ciclo da soja, por meio de pluviômetro digital ou modelo São Izidro, distante no máximo 2 km da lavoura, e os dados de temperatura máxima, mínima e média, coletados pelas estações meteorológicas das fazendas, INMET e Fundações de Pesquisa.

ii. Dados de solos: As coletas de solos foram feitas no momento da colheita da soja. Em cada lavoura foi realizada a coleta composta de solos, sendo de 12 subamostras na camada superficial (0 a 20 cm de profundidade) e 3 subamostras nas camadas abaixo (de 20 cm até 100 cm), dentro da área delimitada de acompanhamento, de 2,5 a 10 hectares. O equipamento utilizado foi o trado holandês, sendo que em cada amostragem foi feita a limpeza do trado para eliminar a contaminação pelas diferentes camadas do solo. As amostras foram estratificadas em 0-10 cm, 10-20 cm, 20-40 cm, 40-60 cm, 60-80 cm, 80-100 cm de profundidade. As análises químicas de solo foram feitas em colaboração com o IBRA com as seguintes análises: pH CaCl, pH KCl, pH água, matéria orgânica (dicromato de potássio), fósforo (g.dm^{-3}), potássio, cálcio e magnésio pelo método da resina (mmolc.dm^{-3}), enxofre pelo método do fosfato de cálcio (g.dm^{-3}), alumínio (Extraído por KCl) e micronutrientes: boro (mg.dm^{-3} , água quente) e zinco, ferro, cobre e manganês (mg.dm^{-3} , DTPA) e argila, silte, areia (metodologia da pipeta), metodologias descritas em Raij et al (1999).

Em cada lavoura foi realizado, ainda, o levantamento da resistência do solo, totalizando 6 a 10 avaliações por lavoura, por meio do penetrômetro modelo HATO, descrito por Paiva de Lima e colaboradores (2013). Todas as avaliações foram feitas com a umidade próxima à capacidade de campo.

iii. Manejo. Foram levantadas as datas de semeadura e colheita, identificando dessa forma seu ciclo, e a distribuição de plantas pelo coeficiente de variação.

iv. Produtividade: A produtividade foi avaliada pelo método gravimétrico. Antes de iniciar o procedimento de colheita das áreas, foi feita a limpeza da colhedora e do caminhão, averiguada a calibração da balança e certificada pelo INMETRO, avaliado o peso total da área colhida e corrigida a umidade dos grãos para 14%, padrão proposto pelo CESB. A área colhida foi medida três vezes com uso de GPS Garmin modelo ETREX3. O tamanho da área colhida foi de 3 hectares.

Os dados foram submetidos à análise de correlação de Pearson e foram destacadas as 20 variáveis significativas com maior associação positiva ou negativa na produtividade. Dessas, foram separadas em produtividade em acima e abaixo da mediana, que permite dividir o banco de dados em dois volumes semelhantes de dados e caracterizados pelo teste t.

Os algoritmos de predição foram construídos pelas plataformas Python (Python Software Foundation, 2009) e SAS (SAS Institute Inc., 2020). Foram aplicadas metodologias de predição quantitativa por meio da Rede Neural e da Floresta Randômica. Foi aplicado nos modelos da Floresta Randômica e da Rede Neural o treinamento

supervisionado, em que as variáveis de entrada do modelo (variáveis agronômicas) são conhecidas e parcialmente controladas. Os modelos foram analisados pelo erro absoluto.

A análise de predição qualitativa foi avaliada pelos modelos de Floresta Randômica, de Análise Discriminante e Suport Vector Machine. A Análise Discriminante teve o processo de Step Wise como procedimento para a construção do modelo, em que consiste numa série de tentativas e avaliações das variáveis, e são selecionadas as variáveis em que se obtém o melhor resultado, as variáveis avaliadas têm como critério um nível de significância de no mínimo 15%. O valor de divisão do banco de dados teve a mediana ou o terceiro quartil como divisor de classificação.

Para a elaboração e treinamento do modelo os dados foram escolhidos aleatoriamente dentre os dados coletados, para fazer a validação do modelo pelo método de CrossValidation, que consiste na retirada aleatória de amostras para avaliar a validade do modelo construído (REFAEILZADEH; TANG; LIU, 2009).

4. RESULTADOS E DISCUSSÃO

4.1. Análise descritiva do banco de dados

Foram reunidos e levantados os dados agronômicos de solo, produtividade e clima de 91 lavouras (Tabela 1). A mediana do banco de dados ficou em 4.920 kg.ha⁻¹, é uma mediana de produtividade alta se considerar que no Brasil temos 2.080 municípios produtores de soja e apenas 9 possuem produtividade acima de 4.200 kg.ha⁻¹ (SAKO et al., 2016). A produtividade máxima do banco de dados ficou em 8.507 kg.ha⁻¹ e o limite superior em 5.185 kg.ha⁻¹.

Tabela 1. Estatística descritiva do banco de dados reunido na dissertação.

Itens	Valores
Unidades de observação	91
Produtividade Amplitude	3.383 a 8.507 kg.ha ⁻¹
Mediana	4.920 kg.ha ⁻¹
Média	4.985 kg.ha ⁻¹
Primeiro Quartil	4.239
Terceiro Quartil	5.500
Limites inferior e superior (teste t a 95%)	4.785 a 5.185 kg.ha ⁻¹

4.2. Fatores climáticos e da fertilidade de solo na produtividade da soja

Foram identificadas 43 das 219 variáveis agronômicas com maior associação positiva ou negativa com a produtividade (20% de todas as variáveis agronômicas). Dessas 43 variáveis agronômicas, 27,90% estão associadas a fatores de clima, 4,65% estão associadas a variáveis do solo muito difíceis de serem alteradas e 67,45% a variáveis de solo que podem ser alteradas por manejo. As variáveis de solo que podem ser alteradas pelo manejo dividem-se entre 23,26% que estão nos horizontes superficiais (0-20 cm de profundidade) e 44,19% que estão localizadas em subsuperfície (abaixo de 20 cm) (Tabela 2). Os resultados da correlação de Pearson estão apresentados na Tabela 3.

Tabela 2. Aspectos sobre as variáveis agronômicas usadas, número de cada uma delas e o grau de influência na produtividade (%).

Variáveis Agronômicas	Número de variáveis	%
Variáveis de clima	12	27,90%
Variáveis de Solo muito difíceis de serem alteradas	2	4,65%
Variáveis de solo que podem ser alteradas pelo manejo em superfície (até 20cm de profundidade)	10	23,26%
Variáveis de solo que podem ser alteradas pelo manejo em subsuperfície (abaixo de 20cm)	19	44,19%
Total	43	100%

Tabela 3. Correlação de Pearson das variáveis com maior influência na produtividade. Legenda da coluna F: (M) variáveis de solo alteráveis pelo manejo a 20 cm; (MP) variáveis manejáveis abaixo de 20 cm de profundidade; (S) variáveis difíceis de serem alteradas por manejo ou não possíveis de serem alteradas; (C) fatores associados ao clima. De todas as variáveis agrônomicas, as 20% de maior relevância em relação à produtividade estão destacadas em negrito. As siglas acompanhadas nas variáveis climáticas representam o período do estágio fenológico.

Variável	Correlação Pearson	F	Variável	Correlação Pearson	F	Variável	Correlação Pearson	F
Resistência do solo 30 cm	0,57594	MP	Argila 80-100 cm	0,22134	S	Temperatura mínima R5-R6	0,18928	C
Resistência do solo 40 cm	0,42773	MP	Mn 40-60 cm	0,24547	MP	Chuva R5-R6	0,18888	C
Resistência do solo 20 cm	0,41508	M	Mn 20-40 cm	0,22115	MP	Mn 80-100 cm	0,18716	M
Fósforo (P) 0-10 cm	0,3397	M	m (%) 40-60 cm	0,22042	MP	Cu 60-80 cm	0,18609	M
Mg na CTC (%) 0-10 cm	0,3356	M	m (%) 20-40 cm	0,22003	MP	Delta pH 20-40 cm	0,18107	S
pH CaCl ₂ 20-40 cm	0,33485	MP	pH CaCl₂ 60-80 cm	0,21273	MP	% K na CTC 40-60 cm	0,18094	M
DeltapH 80-100 cm	0,32999	S	Chuva Semeadura R8	0,20844	C	Zn 60-80 cm	0,17375	M
pH CaCl ₂ 10-20 cm	0,32586	M	Mn 60-80 cm	0,20572	MP	Déficit hídrico semente a R8	0,17149	C
Enxofre (S-SO ₄) 80-100 cm	0,32279	MP	Delta pH 10-20 cm	0,20163	M	Excedente R5-R6	0,17137	C
Enxofre (S-SO ₄) 60-80 cm	0,3144	MP	Cu 10 a 20 cm	0,20123	M	K (%) 80-100 cm	0,17133	M
Temperatura máxima - R1	0,30849	C	P 10-20 cm	0,19761	M	Delta pH 40-60 cm	0,17018	S
Resistência do solo 50 cm	0,27926	MP	Excedente hídrico semente-R8	0,19737	C	Temperatura mínima semente emergência	0,16849	C
Evapotranspiração R1	0,27672	C	Déficit hídrico semente-R8	0,1964	C	Ca 10-20 cm	0,16786	M
Temperatura máxima R5-R6	0,2663	C	Cu 80-100 cm	0,19343	MP	Argila 0-10 cm	0,1662	S
Mg na CTC (%) 80-100cm	0,25967	MP	m (%) 10-20 cm	0,19224	M	Evapotranspiração R1 a R5	0,16619	C
Cobre (Cu) 40-60cm	0,25425	MP	m (%) 80-100 cm	0,18989	MP	Déficit hídrico semente-R8	0,15393	C
pH CaCl ₂ 80-100 cm	0,25221	MP	Temperatura base R5-R6	0,18928	C	Temperatura média semente-R8	0,15048	C
Cobre (Cu) 20-40 cm	0,25129	MP	Chuva R5-R6	0,18888	M	Evapotranspiração R5-R6	0,15031	C
Temperat. máxima semente-emerg.	0,24915	C	Mn 80-100 cm	0,18716	C	m (%) 60-80 cm	0,14924	M
Enxofre (S-SO ₄) 40-60 cm	0,24776	MP	Cu 60-80 cm	0,18609	C	Argila 20-40 cm	0,14789	S
pH CaCl ₂ 0-10 cm	0,24639	M	Delta pH 20-40cm	0,18107	M	Ca 0-10 cm	0,14768	M
Manganês (Mn) 40-60 cm	0,24547	MP	Mn 10-20 cm	0,19624	M	P 60-80 cm	0,14597	S
Temperatura média R5-R6	0,23876	C	Cu 0-10 cm	0,19574	M	CTC 10-20 cm	0,14437	M
Temperatura média emergencia-R1	0,23736	C	Mn 0-10 cm	0,19436	M	% K na CTC 0-10 cm	0,14404	M
Temperatura mínima R7-R8	0,23608	C	Cu 80-100 cm	0,19343	M			
Saturação de Al (m%) 0-10 cm	0,23047	M	m (%) 1 10-20 cm	0,19224	M			

Para caracterizar a fertilidade de superfície e de subsuperfície e a qualidade física do solo nas produtividades abaixo e acima da mediana, $4.920 \text{ kg}\cdot\text{ha}^{-1}$, foi aplicado o teste t e apresentado nas Tabelas 4 a 9. Uma vez que são poucos os trabalhos que retratam tecnicamente a alta produtividade de soja, é importante identificar os valores da fertilidade do solo em que ocorrem as altas produtividades pois esta informação traz reflexões importantes para a pesquisa agrônômica e referências importantes para o manejo. Está apresentado pelo teste t todas as variáveis do solo separadas que representam a produtividade acima e abaixo da mediana e segue a análise da causalidade de algumas das 20% das variáveis agrônômicas com maior associação positiva ou negativa com a produtividade pela correlação de Pearson.

Resistência do solo (30, 40, 50 cm de profundidade): As primeiras variáveis com maior associação com a produtividade foram a resistência do solo (Figuras 4 a 7). Para 1 MPa que se reduziu houve um acréscimo de $1.200 \text{ kg}\cdot\text{ha}^{-1}$ na camada de 30cm de profundidade. Sako et al. (2016), realizando um levantamento semelhante em diversas lavouras, identificaram 5 fatores relevantes para produtividades acima de $4.200 \text{ kg}\cdot\text{ha}^{-1}$, dentre eles, a resistência de solo. Busscher et al. (2000) identificaram que para cada 1 MPa reduzido houve um acréscimo de $720 \text{ kg}\cdot\text{ha}^{-1}$. Levantamentos extensos feitos pelo CESB desde 2014 em campeões de produtividade demonstram valores menores que 1,7 MPa (SAKO, 2019). De outro aspecto, Avelar et al. (2014) identificaram que valores de resistência de solo extremamente baixos provocaram queda na produtividade, nesse caso, devido à formação de bolsões de ar por subsolagem em umidade fora da recomendada. No presente estudo, os valores de resistência do solo estiveram abaixo de 1,5 MPa para altas produtividades de soja (Tabela 4).

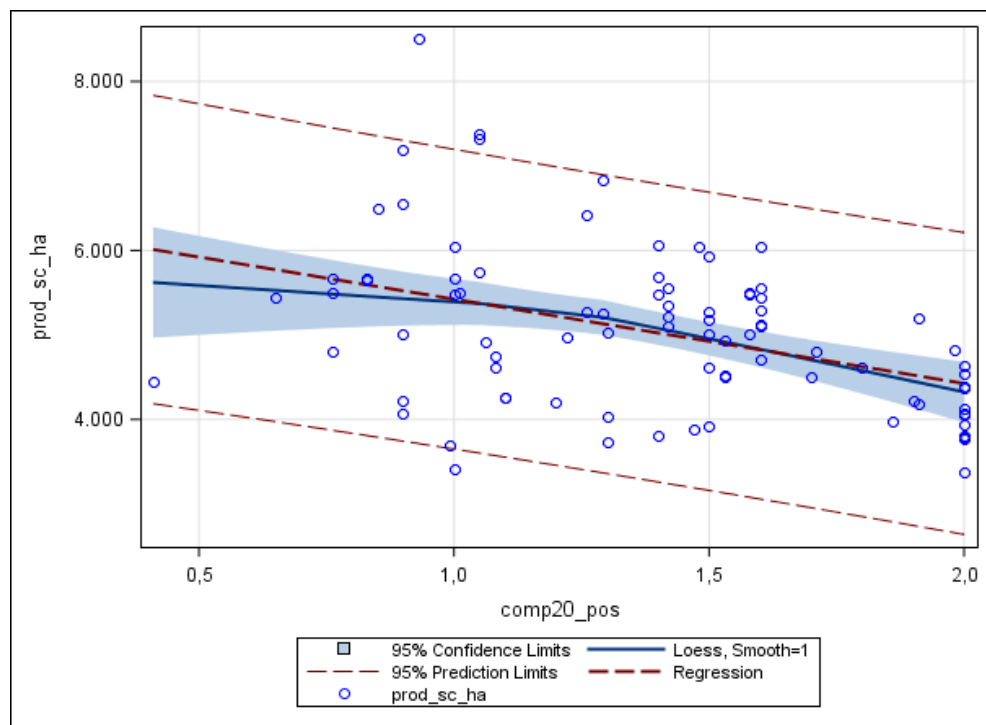


Figura 4. Valores de resistência do solo a 20 cm de profundidade (MPa) no eixo x e de produtividade ($\text{kg}\cdot\text{ha}^{-1}$) no eixo y.

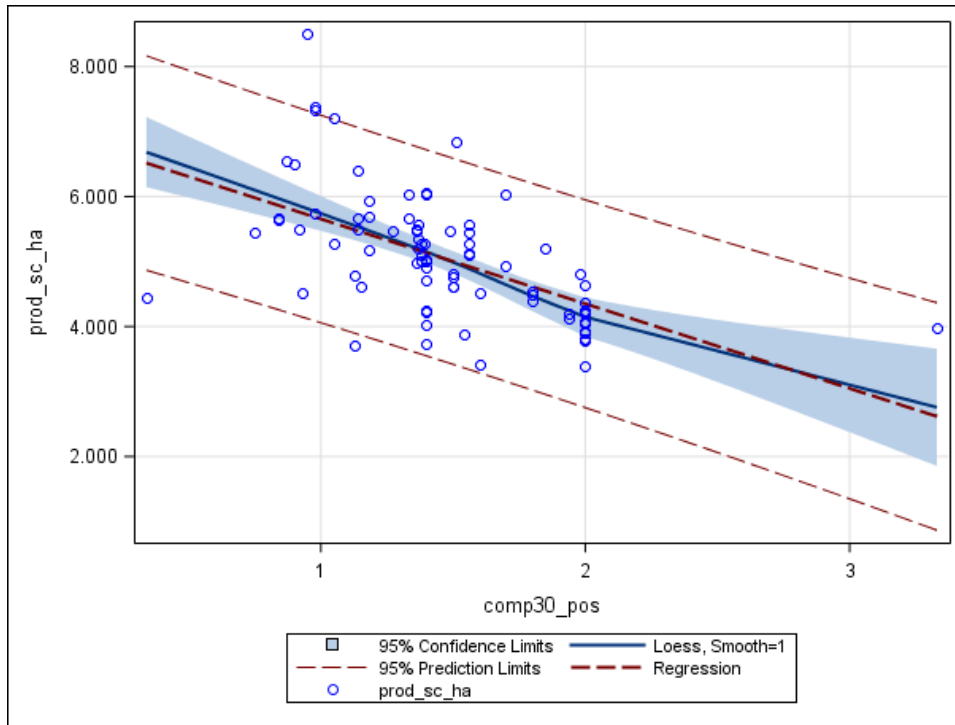


Figura 5. Valores de resistência do solo a 30 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha⁻¹) no eixo y.

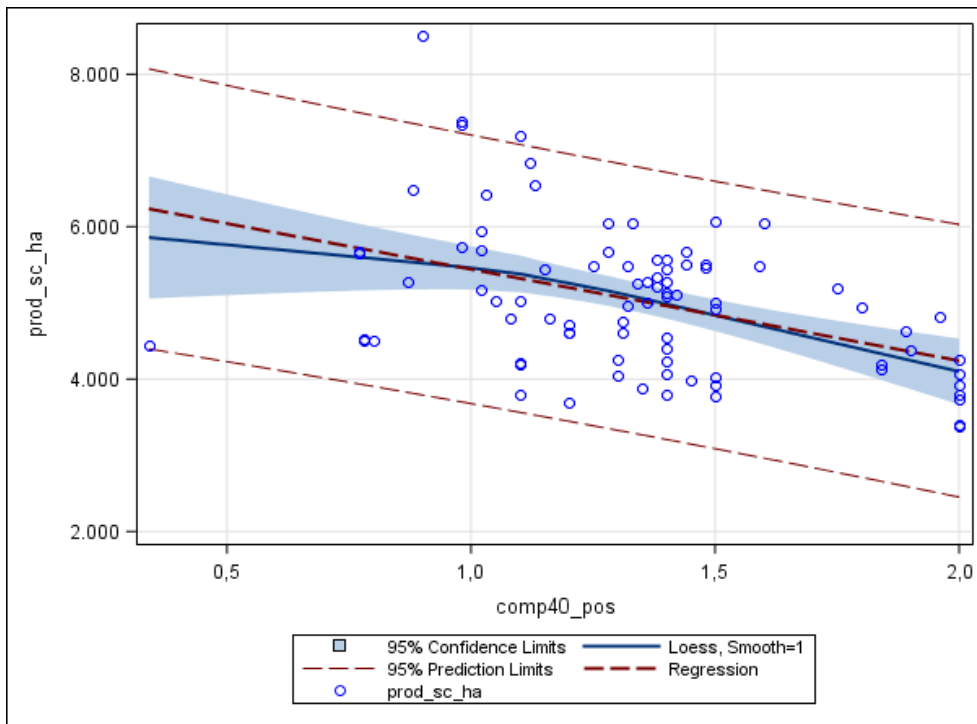


Figura 6. Valores de resistência do solo a 40 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha⁻¹) no eixo y.

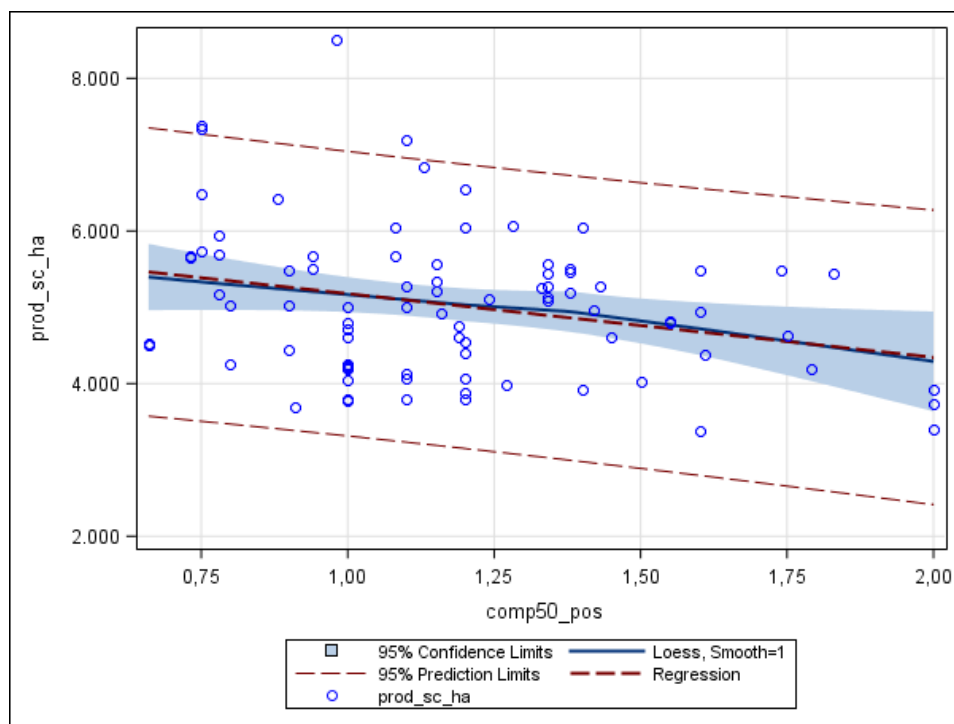


Figura 7. Valores de resistência do solo a 50 cm de profundidade (MPa) no eixo x e de produtividade (kg.ha⁻¹) no eixo y.

Tabela 4. Valores de resistência do solo (Res) em dois grupos de produtividades de soja: abaixo e acima de 4.920 kg.ha⁻¹.

Variável	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg.ha ⁻¹	
	Límite Inferior	Límite Superior	Límite Inferior	Límite Superior
Res 0-10 cm	0,9	1,1	0,9	1,2
Res 10-20 cm	1,1	1,3	1,3	1,6
Res 20-30 cm	1,1	1,3	1,5	1,8
Res 30-40 cm	1,1	1,3	1,3	1,5
Res 40-50 cm	1,0	1,2	1,1	1,3

Teores de Fósforo (0-10 cm, 10-20 cm de profundidade): A ciência de solos tem obtido grandes avanços no estudo dos solos tropicais e uma referência importante para todos os técnicos de classificação da fertilidade do solo é o boletim 100, publicado em 1997. No entanto, a produtividade de soja dessa época e seus materiais genéticos são distintos do atual período, causando sempre uma reflexão se essas referências podem ou não ter mudado. Os teores médios no boletim 100 definem a resposta de 90% da máxima produtividade quando o fósforo é de 16 a 40 mg.dm⁻³ e no presente estudo foi identificado de 57 a 81 mg.dm⁻³ na camada 0-10 cm e na camada de 10-20 cm de 36 a 60 mg.dm⁻³, como mostrado nas Figuras 8 e 9, a relação do fósforo com a produtividade da soja. De modo semelhante, os teores de potássio, magnésio, cálcio, boro, manganês, zinco e cobre apresentam valores superiores aos sugeridos pelo boletim 100 (Tabelas 6, 7, 8 e 9). Sugerem-se maiores estudos nesse contexto, buscando avaliar se os teores médios de macronutrientes e micronutrientes descritos no boletim 100 ainda representam a resposta de 90% da máxima produtividade no atual sistema de produção.

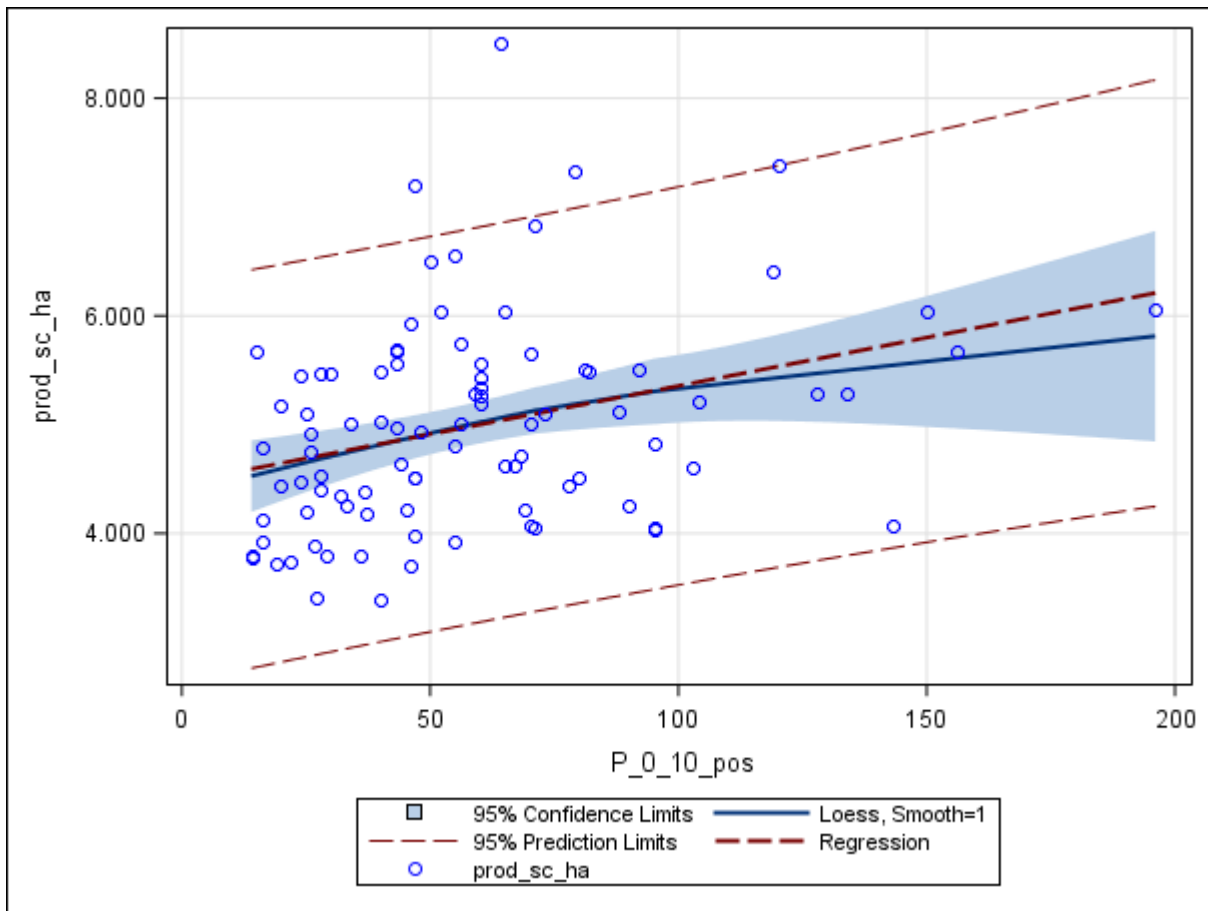


Figura 8. Teores de fósforo (mg.dm⁻³, Resina) na camada 0-10 cm no eixo x e produtividade da soja (kg.ha⁻¹) no eixo y.

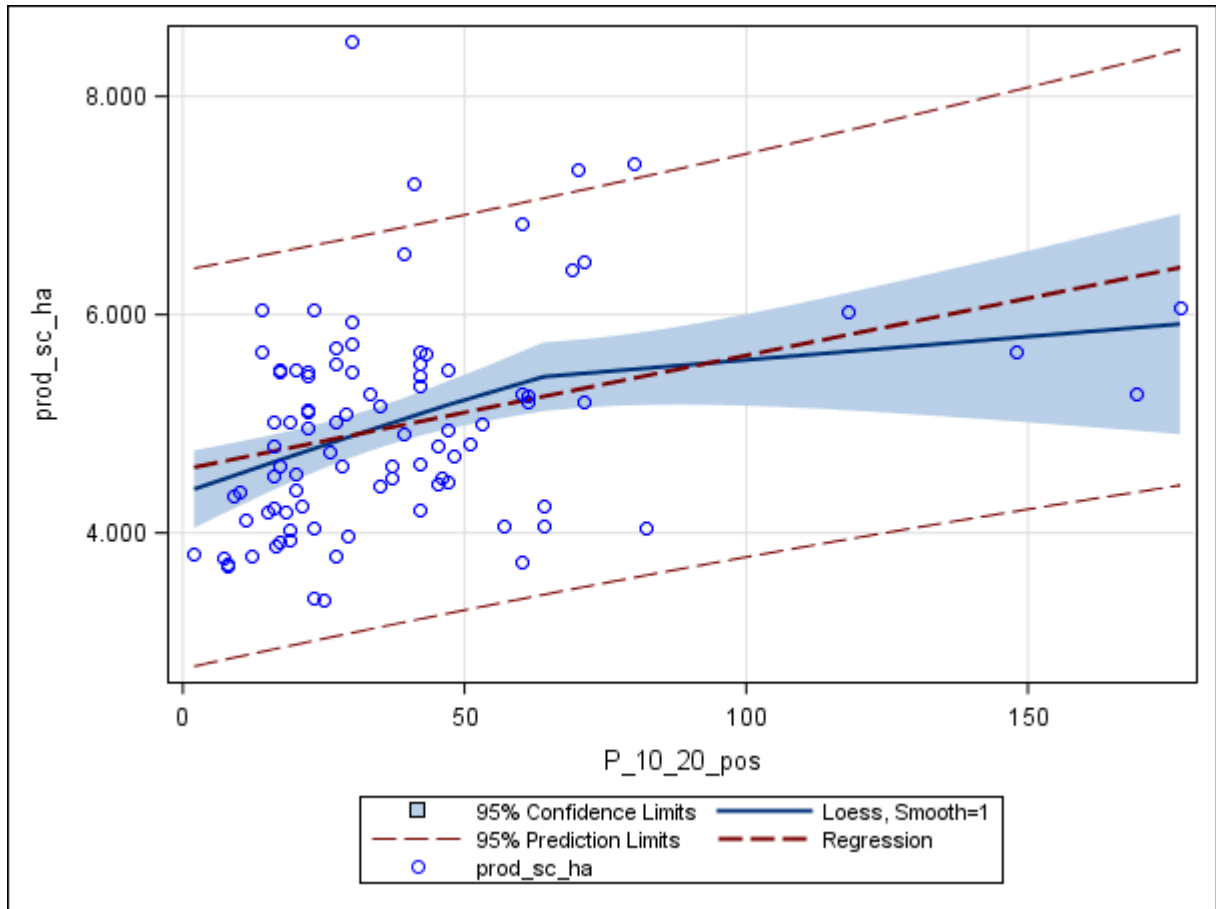


Figura 9. Teores de fósforo ($\text{mg}\cdot\text{dm}^{-3}$, Resina) na camada 10-20 cm no eixo x e produtividade da soja ($\text{kg}\cdot\text{ha}^{-1}$) no eixo y.

Tabela 5. Teores de fósforo (P, mg.dm⁻³), potássio (K, mmolc.dm⁻³), magnésio (Mg, mmolc.dm⁻³), cálcio (Ca, mmolc.dm⁻³) em produtividades acima e abaixo de 4.920 kg.ha⁻¹, teste t.

Teores de nutriente no solo e sua profundidade	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg. ha ⁻¹	
	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
P 0-10 cm	57,3	81,1	39,5	56,4
P 10-20 cm	36,7	60,1	24,2	34,8
P 20-40 cm	11,7	34,3	8,8	13,9
P 40-60 cm	4,0	8,6	3,6	10,3
P 60-80 cm	3,0	5,3	3,3	7,8
P 80-100 cm	2,9	4,8	3,7	5,9
K 0-10 cm	2,6	3,9	2,5	3,5
K 10-20 cm	2,0	2,7	1,8	2,6
K 20-40 cm	1,3	1,8	1,1	1,8
K 40-60 cm	0,9	1,2	0,8	1,3
K 60-80 cm	0,6	1,0	0,6	1,1
K 80-100 cm	0,4	0,8	0,5	0,8
Mg 0-10 cm	12,3	17,0	12,9	17,0
Mg 10-20 cm	9,8	13,8	9,8	13,5
Mg 20-40 cm	7,2	10,8	6,8	10,2
Mg 40-60 cm	5,7	8,0	5,4	8,3
Mg 60-80 cm	5,0	7,6	4,7	7,3
Mg 80-100 cm	4,6	6,8	4,2	6,2
Ca 0-10 cm	33,5	45,3	35,3	44,1
Ca 10-20 cm	30,0	38,5	27,8	36,8
Ca 20-40 cm	21,1	29,4	18,7	28,7
Ca 40-60 cm	13,3	18,5	13,0	21,7
Ca 60-80 cm	11,9	16,9	11,6	20,5
Ca 80-100 cm	10,5	15,3	10,9	19,8

Acidez (pH CaCl₂ a 10-20 cm, 20-40 cm, 60-80 cm, 80-100 cm de profundidade): A acidez é um atributo relevante nos solos tropicais pois influencia diversos processos no ambiente de produção como: i) atividade do nódulo: a atividade dos nódulos é diretamente influenciada pelo pH. Quando o solo apresentou pH maior que 5,1 o nitrogênio fixado foi de 335 kg.ha⁻¹ e quando o pH foi menor que 5,1 o nitrogênio fixado variou entre 129 a 223 kg.ha⁻¹, dependendo da saturação de alumínio (ALVES et al., 2020). É importante ressaltar que para boas produtividades um grande desafio é o atendimento de nitrogênio via fixação biológica de nitrogênio (CIAMPITTI; SALVAGIOTTI, 2018). ii) fixação de fósforo: a acidez reduz a fixação de fósforo pelos óxidos de ferro e alumínio (SANCHEZ, 2019). iii) correção do alumínio tóxico: o aumento do pH corrige o alumínio tóxico (SANCHEZ, 2019). iv) correção do perfil do solo: a preservação de maiores valores de pH nas primeiras camadas do solo constrói uma frente alcalinizante que é um primeiro passo para a correção do perfil do solo (CAIRES et al., 2008).

As práticas agrícolas causam constantemente acidez no solo, seja pelos adubos nitrogenados, chuva, erosão laminar, exportação de nutrientes e, principalmente, pela própria atividade do nódulo (NYATSANGA; PIERRE, 1973; ROBSON, 1989). Pode-se considerar, então, que a acidez é o “Imposto Biológico da Agricultura, produzir é gerar acidez”. Segue nas Figuras 10 a 14 a relação da acidez em diversas profundidades com a produtividade da soja e os valores de acidez identificados na produtividade acima da mediana da soja pelo teste t na Tabela 6. É notável o aumento do pH associado ao aumento da produtividade em todo o perfil do solo e vale destacar a associação positiva do aumento do pH com a produtividade nas camadas do solo abaixo de 20 cm de profundidade.

Tabela 6. Valores de pH CaCl₂ em produtividades acima e abaixo de 4.920 kg.ha⁻¹.

Variável	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg.ha ⁻¹	
	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
pH 0-10 cm	4,9	5,2	4,9	5,1
pH 10-20 cm	4,9	5,1	4,8	5,0
pH 20-40 cm	4,8	5,0	4,7	4,9
pH 40-60 cm	4,5	5,0	4,8	4,9
pH 60-80 cm	4,9	5,0	4,7	4,9
pH 80-100 cm	4,9	5,0	4,8	5,0

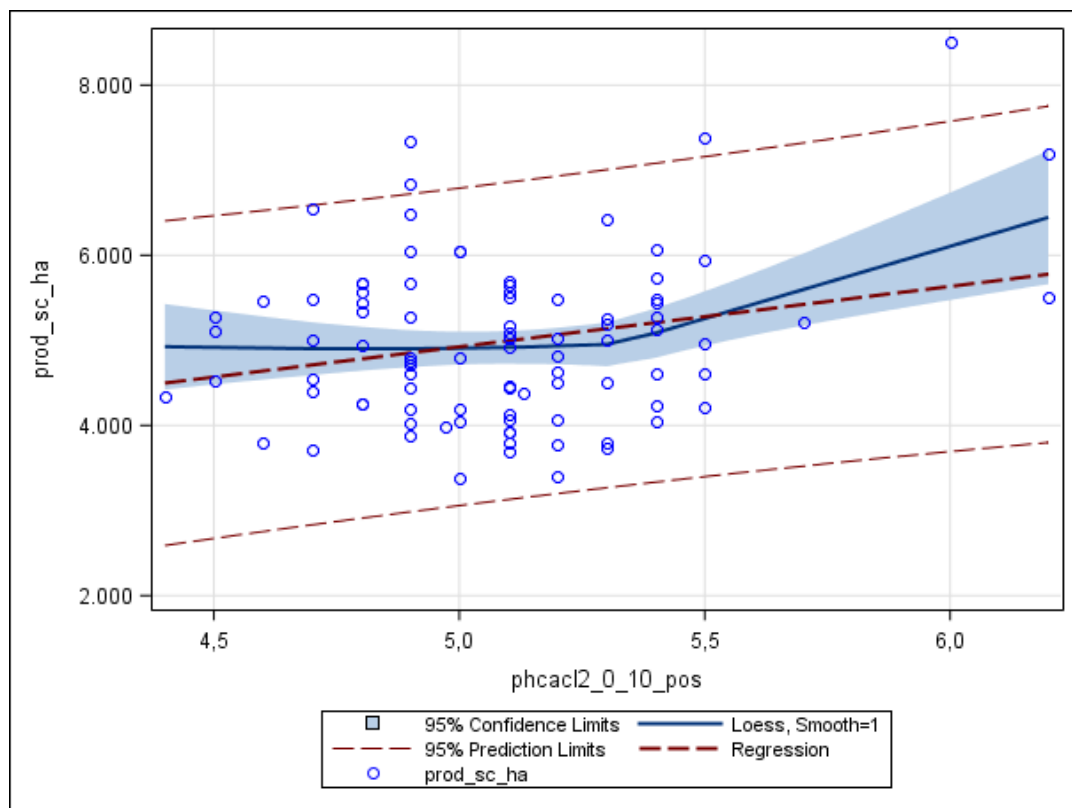


Figura 10. Relação entre o pH CaCl₂ na camada de 0-10 cm de profundidade e a produtividade de soja (kg.ha⁻¹).

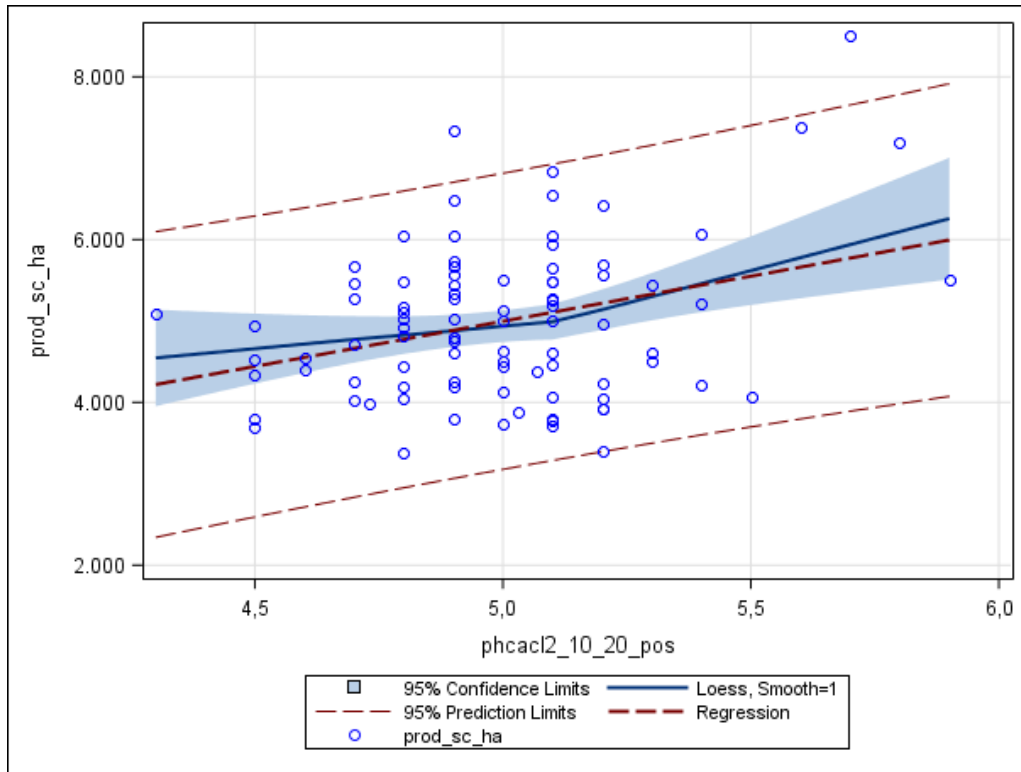


Figura 11. Relação entre o pH CaCl_2 na camada de 10-20 cm de profundidade e a produtividade de soja ($\text{kg}\cdot\text{ha}^{-1}$).

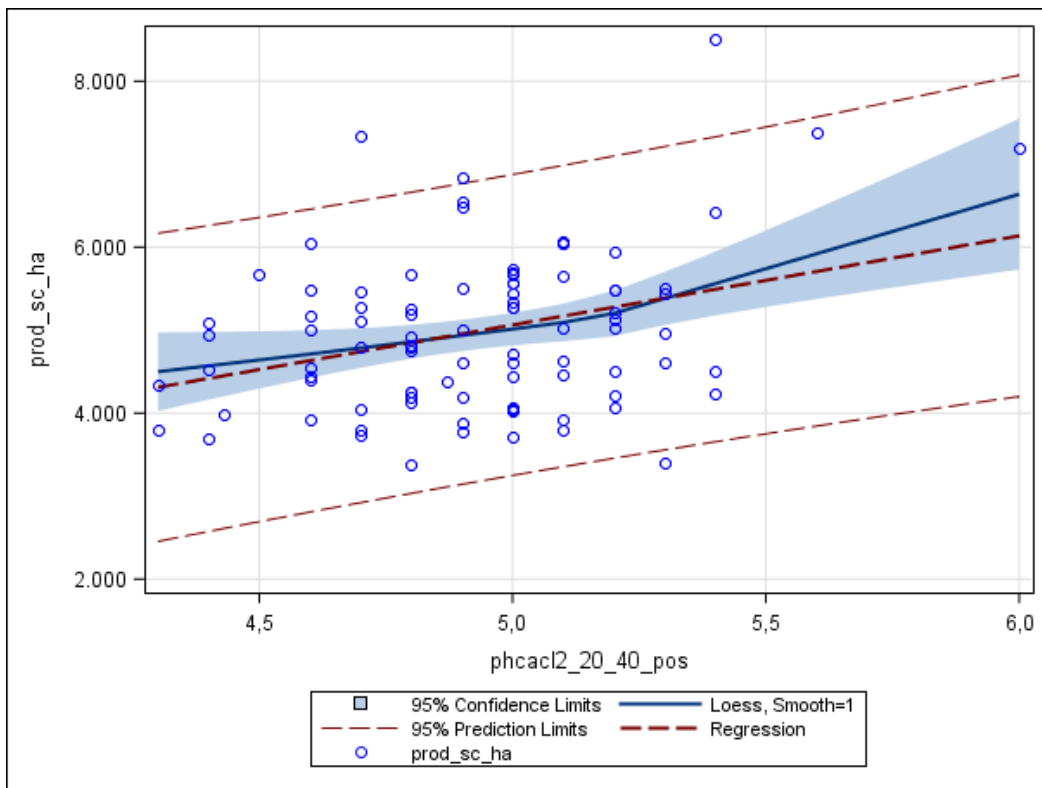


Figura 12. Relação entre o pH CaCl_2 na camada de 20-40 cm de profundidade e a produtividade da soja ($\text{kg}\cdot\text{ha}^{-1}$).

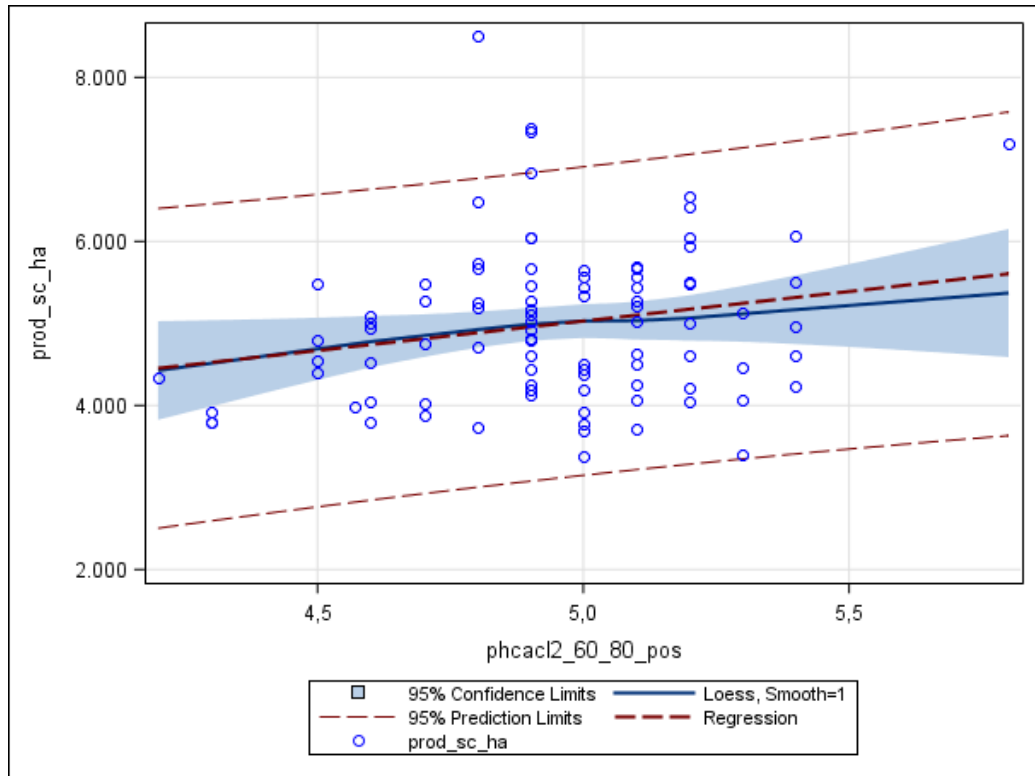


Figura 13. Relação entre o pH CaCl_2 a 60-80 cm de profundidade e a produtividade da soja ($\text{kg}\cdot\text{ha}^{-1}$).

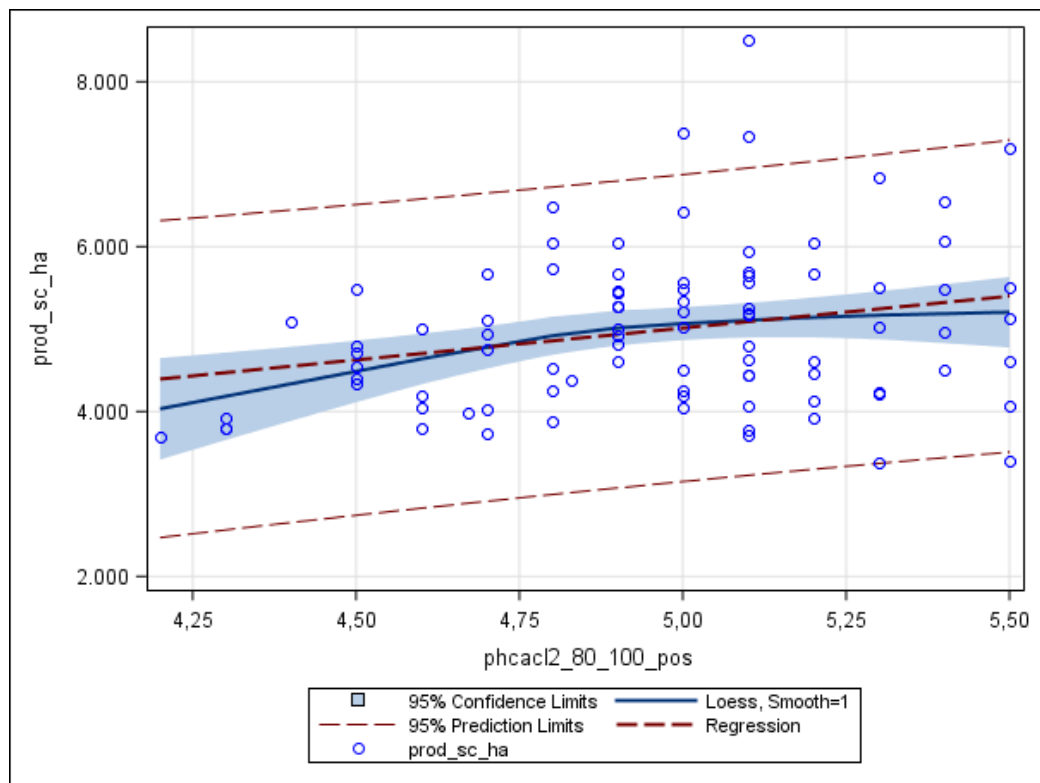


Figura 14. Relação entre o pH CaCl_2 a 80-100 cm de profundidade e a produtividade da soja ($\text{kg}\cdot\text{ha}^{-1}$).

Participação do Magnésio na CTC (0-10 cm, 80-100 cm): A proporção dos cátions K, Mg e Ca é citada frequentemente nos trabalhos e há preocupação entre os técnicos atualmente em buscar seu equilíbrio. Dentre eles, vale destacar o magnésio: o magnésio possui um papel importante na distribuição do carboidrato das folhas para as raízes. Marschner (2011), avaliando a deficiência de magnésio, notaram que a concentração de carboidrato nas raízes era de 0,8% do total produzido, enquanto que a deficiência de potássio levou ao acúmulo de 3,4%, e as plantas bem nutridas tinham 15,7%. Plantas bem supridas de carboidratos possuem maior capacidade de tolerar o calor devido à formação de raízes e à melhor conversão da energia solar em carboidratos, reduzindo a energia livre no dossel das folhas. Dentre os três nutrientes K, Mg e Ca, o magnésio é o elemento que não possui um canal específico de absorção de nutrientes, ficando sempre prejudicado entre ele e o potássio, que tem grande facilidade de ser absorvido pela planta (NARWAL; KUMAR; SINGH, 1985). Aplicando doses crescentes de magnésio no solo, Narwal e colaboradores (1985) não notaram decréscimo nos teores de potássio na folha e na raiz, já o aumento dos teores de potássio levou ao decréscimo de magnésio na folha e na raiz. Por esse motivo, as aplicações de potássio e a correção de magnésio devem ser bem ponderados. Os valores de porcentagem dos nutrientes K, Mg e Ca auferidos em diversas profundidades no banco de dados avaliado neste trabalho seguem na Tabela 7.

Tabela 7. Porcentagem de potássio, magnésio e cálcio na CTC em produtividades acima e abaixo de 4.920 kg.ha⁻¹.

Nutriente no solo e sua profundidade	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg.ha ⁻¹	
	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
%K 0-10 cm	3,1	4,2	2,9	3,7
%K 10-20 cm	2,5	3,4	2,3	3,1
%K 20-40 cm	2,1	2,8	1,8	2,6
%K 40-60 cm	1,7	2,3	1,5	2,1
%K 60-80 cm	1,5	2,1	1,3	1,9
%K 80-100 cm	1,2	1,9	1,0	1,6
%Mg 0-10 cm	14,4	18,0	15,3	17,8
%Mg 10-20 cm	12,2	15,6	12,9	15,9
%Mg 20-40 cm	10,6	13,6	10,7	13,6
%Mg 40-60 cm	10,5	13,0	10,1	12,7
%Mg 60-80 cm	10,4	12,7	9,7	12,3
%Mg 80-100 cm	10,4	13,0	9,2	11,5
%Ca 0-10 cm	42,5	46,6	42,1	46,8
%Ca 10-20 cm	38,1	43,7	36,8	42,7
%Ca 20-40 cm	32,2	37,9	29,4	37,3
%Ca 40-60 cm	25,1	30,7	24,0	31,9
%Ca 60-80 cm	24,6	30,0	23,7	31,8
%Ca 80-100 cm	23,2	29,1	22,8	31,3

Saturação de alumínio (0-10 cm, 10-20 cm, 20-40 cm, 40-60 cm, 80-100 cm de profundidade): Os solos tropicais são ácidos ao longo do perfil do solo e a acidez ativa gerada pelo alumínio tóxico penaliza o crescimento radicular e, conseqüentemente, a produtividade (SANCHEZ, 2019). O alumínio causa a desorganização da divisão e crescimento das células radiculares por meio da desorganização dos microtúbulos e microfilamentos (SHIVAGURU et al., 1999; MATSUMOTO, 2000; SIMONOVICOVA et al., 2004), o alumínio também reduz a

permeabilidade da membrana celular (MATSUMOTO, 2000), e reduz a atividade dos nódulos da soja (SARTAIN, KAMPRATH, 1977) e o mesmo pode ser absorvido e translocado para a parte aérea, danificando os componentes do aparato fotossintético.

Os teores dos nutrientes S, B, Cu, Zn, Mn e Ferro, de MO e as saturações por bases e de alumínio nas profundidades estudadas para as áreas consideradas neste estudo encontram-se nas Tabelas 8 e 9.

Tabela 8. Fertilidade do perfil de solo entre produtividades acima e abaixo de 4.920 kg.ha⁻¹, os valores foram delimitados pelo teste t a 5% de significância para saturação por bases (V%) enxofre (S), matéria orgânica (MO) e saturação de alumínio (m%).

Teores de nutriente no solo e sua profundidade	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg.ha ⁻¹	
	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
V% 0-10 cm	61,1	67,8	61,8	68,3
V% 10-20 cm	54,2	62,1	53,0	61,3
V% 20-40 cm	45,9	54,0	42,4	52,8
V% 40-60 cm	38,5	46,4	38,1	48,1
V% 60-80 cm	37,8	45,2	35,3	45,4
V% 80-100 cm	37,9	43,7	33,7	43,1
S 0-10 cm	14,9	19,7	8,5	24,9
S 10-20 cm	17,1	24,7	14,2	22,2
S 20-40 cm	28,5	42,0	23,5	33,4
S 40-60 cm	39,5	66,3	30,3	43,7
S 60-80 cm	37,5	72,2	24,5	34,4
S 80-100 cm	30,7	69,1	22,5	33,3
MO 0-10 cm	28,7	35,1	29,4	35,8
MO 10-20 cm	24,9	30,7	24,6	30,9
MO 20-40 cm	19,7	25,3	19,5	24,9
MO 40-60cm	14,7	19,2	14,9	19,6
MO 60-80 cm	13,4	18,1	13,3	18,3
MO 80-100 cm	11,9	15,7	11,6	15,8
m% 0-10 cm	0,0	0,2	0,0	0,8
m% 10-20 cm	0,0	3,1	1,0	6,2
m% 20-40 cm	0,7	5,7	4,4	11,6
m% 40-60 cm	2,0	7,4	5,7	15,2
m% 60-80 cm	1,4	7,5	3,8	14,9
m% 80-100 cm	0,0	5,5	3,0	13,6

Tabela 9. Valores de nutrientes estabelecidos pelo teste t em produtividades abaixo e acima de 4.920 kg.ha⁻¹. Boro (B); Cobre (Cu) Zinco (Zn) Manganês (Mn) Ferro (Fe).

Teores de nutriente no solo e sua profundidade	Produtividade acima de 4.920 kg.ha ⁻¹		Produtividade abaixo de 4.920 kg.ha ⁻¹	
	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
B 0-10 cm	0,6	1,2	0,6	1,3
B 10-20 cm	0,5	1,1	0,5	0,7
B 20-40 cm	0,4	0,7	0,4	0,6
B 40-60 cm	0,3	0,6	0,3	0,5
B 60-80 cm	0,3	0,5	0,3	0,5
B 80-100 cm	0,3	0,5	0,3	0,5
Mn 0-10 cm	4,3	22,7	8,9	26,3
Mn 10-20 cm	2,8	22,6	8,1	29,0
Mn 20-40 cm	1,7	11,0	4,0	16,4
Mn 40-60 cm	1,0	3,3	1,5	8,4
Mn 60-80 cm	1,1	3,2	1,2	5,1
Mn 80-100 cm	0,7	2,0	0,6	3,6
Zn 0-10 cm	2,8	7,4	2,7	6,0
Zn 10-20 cm	1,8	6,9	1,7	3,3
Zn 20-40 cm	0,6	1,5	0,6	1,6
Zn 40-60 cm	0,1	0,9	0,3	0,8
Zn 60-80 cm	0,2	0,4	0,2	0,8
Zn 80-100 cm	0,1	0,6	0,2	0,6
Fe 0-10 cm	19,5	31,9	22,6	38,2
Fe 10-20 cm	19,9	28,3	23,4	38,3
Fe 20-40 cm	15,3	19,0	15,2	25,9
Fe 40-60 cm	7,7	10,5	8,8	14,9
Fe 60-80 cm	5,6	8,9	6,5	12,7
Fe 80-100 cm	4,3	7,2	5,2	9,8
Cu 0-10 cm	1,5	2,7	1,9	3,9
Cu 10-20 cm	1,3	2,4	1,6	3,7
Cu 20-40 cm	0,9	1,6	1,2	3,1
Cu 40-60 cm	0,4	0,9	0,8	1,7
Cu 60-80 cm	0,4	0,8	0,5	1,2
Cu 80-100 cm	0,2	0,5	0,4	1,1

A fertilidade de subsuperfície do solo possui uma influência importante na produtividade da soja. Nesse trabalho, a partir de uma relação ampla de variáveis agrônômicas, 40% dos fatores mais relevantes e significativos para a produtividade da soja estão em fertilidade da subsuperfície, ou seja, nos dados coletados de 20 cm até um metro de profundidade.

É, portanto, reflexão importante, a inclusão do levantamento da fertilidade em subsuperfície no sistema grãos para o entendimento do ambiente solo e seu potencial produtivo. Nas Tabelas 4, 5, 6, 7, 8 e 9, por meio da aplicação do teste t, está descrito como é o retrato da fertilidade em subsuperfície das lavouras de produtividade acima da mediana. Dentre as variáveis da fertilidade de subsuperfície, destacam-se alumínio, pH, cálcio, magnésio,

cobre, manganês e zinco. A implicação prática dessa referência técnica é mostrar a importância do levantamento da fertilidade de subsuperfície para entender produtividades acima de 4.920 kg.ha⁻¹.

4.3. Erro Absoluto dos modelos quantitativos de predição

A regressão linear apresentou um erro absoluto de 801,6 kg.ha⁻¹ de soja, a Rede Neural de 655,2 kg.ha⁻¹ e a Floresta Randômica de 568,2 kg.ha⁻¹, ou seja, a Floresta Randômica apresentou o menor erro absoluto (Tabela 10). Apesar de a Floresta Randômica ter construído o melhor modelo, ainda são valores de erro absoluto altos quando comparados com outros trabalhos. Kaul et al. (2005) construíram modelos de predição de produtividade de soja com Rede Neural a níveis regionais com erro absoluto nos melhores locais entre 140 a 195 kg.ha⁻¹, dependendo do número de nós da Rede Neural, e com um extenso banco de dados, levantados de 1978 a 1998 com dados de solo, clima e produtividade de soja. Apesar de ter levantado variáveis agronômicas importantes para a construção do modelo, os valores altos de erro absoluto sugerem a necessidade de um maior número de amostras para aplicar essa metodologia.

Tabela 10. Erro absoluto médio do modelo avaliado.

Modelo avaliado	Erro absoluto	Desvio Padrão
Regressão Linear	801,6	155,64
Rede Neural	655,2	145,62
Floresta Randômica	568,2	95,16

4.4. Modelos de classificação

Os modelos de classificação tiveram resultados melhores de predição que os de regressão. Tendo a mediana do banco de dados como referência de classificação, 4.920 kg.ha⁻¹, a metodologia com melhor índice de acerto foi a Análise Discriminante, com 69% de índice de acerto, seguida da Floresta Randômica, 67%, e Support Vector Machine com 61% (Tabela 11). Com os resultados dos modelos, foi decidido avançar nos estudos com a análise discriminante, por ter a vantagem de sua simplicidade de cálculo, clareza de como é feita a predição e, principalmente, o entendimento da necessidade de um maior número de amostras para poder aplicar as metodologias como a Floresta Randômica e a Support Vector Machine.

Tabela 11. Porcentagem de acerto em classificação tendo o valor da mediana, 4.920 kg.ha⁻¹ como classificador dos valores.

Modelo avaliado	Índice de acerto
Análise Discriminante	69%
Floresta randômica	67%
Support Vector Machine	61%

No modelo da Análise Discriminante, tendo a mediana 4.920 kg.ha⁻¹ como divisor do banco de dados, pelo método StepWise com 15% de significância, foram selecionadas 7 variáveis agronômicas, conforme segue na Tabela 12. O banco de dados possui 91 repetições e 219 variáveis agronômicas, o que torna importante reduzir o número de variáveis para poder restringir as combinações de variáveis agronômicas e com o volume de repetições

conseguir abranger o maior número possível de combinações, de modo que o modelo possa calcular essas possibilidades.

Tabela 12. Variáveis selecionadas no modelo da Análise Discriminante, tendo a mediana como divisor do banco de dados, no valor de 4.950 kg.ha⁻¹.

Variável agrônômica	Produtividade Acima da Mediana	Produtividade Abaixo da Mediana
Constante	-229.08659	-243.24535
pH CaCl ₂ 0-10 cm	25,61418	23,94780
Fósforo 0-10 cm	0,01168	0,01036
Saturação de alumínio 80-100 cm	0,27589	0,34156
Manganês 20-40 cm	-0.14724	-0.18717
Cobre 40-60 cm	1.32046	2.10262
Resistência Solo 30 cm	8.51132	11.70919
Temperatura máxima R5-R6	10.79428	11.41494

A análise canônica dos componentes 1 e 2 gerada pela Análise Discriminante demonstra que o modelo conseguiu classificar razoavelmente as produtividades acima e abaixo da mediana, mas nota-se que o componente 1 teve valores classificatórios muito próximos entre as duas categorias de produtividade, sugerindo que o modelo classifica por pequenas diferenças de valores (Figura 15). Valores diferentes de classificação da produtividade acima e abaixo da mediana podem distinguir melhor os dois grupos e, dessa forma, foi feita a aplicação com o terceiro quartil, 5.500 kg.ha⁻¹.

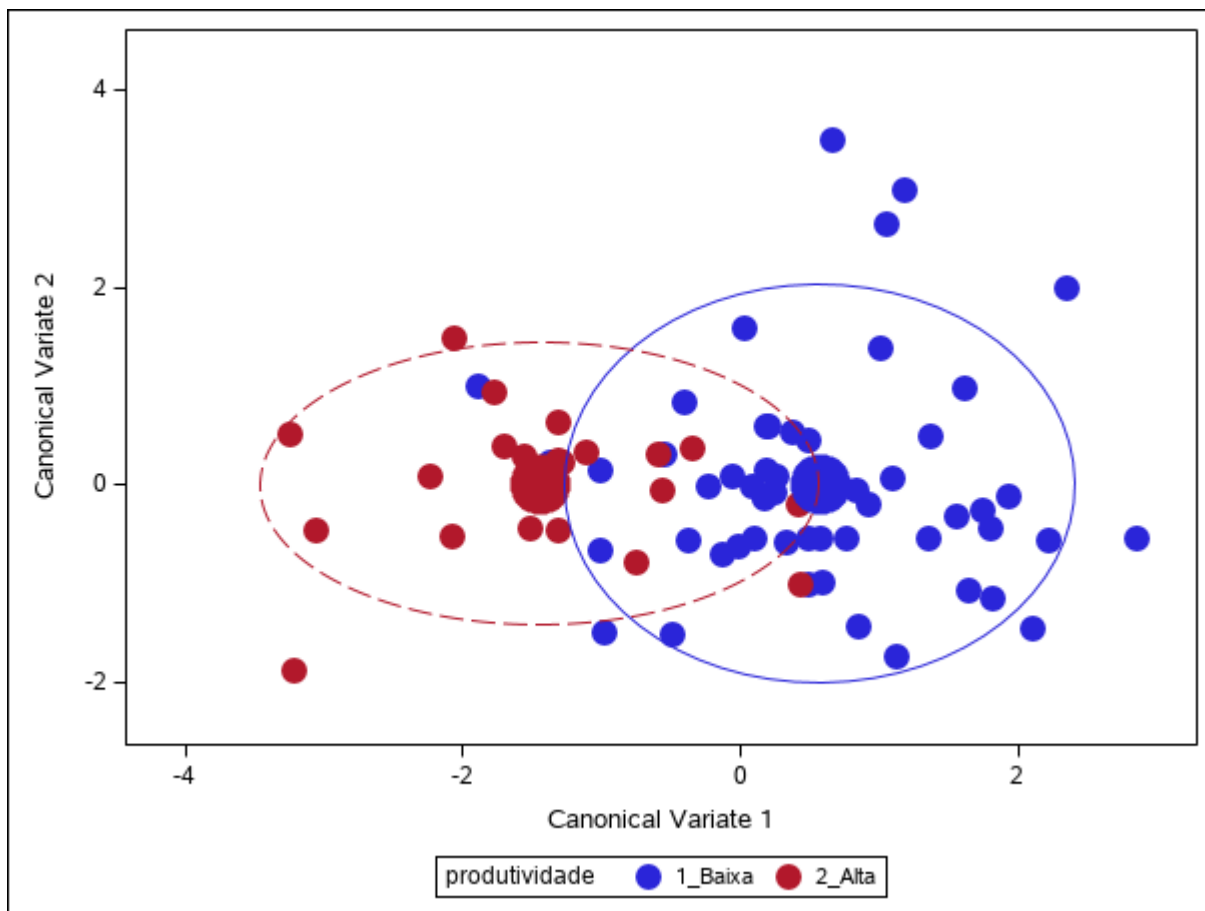


Figura 15. Variáveis canônicas 1 e 2 e a predição da produtividade acima e abaixo da mediana, usando a mediana como divisor do banco de dados.

A Análise Discriminante tendo como divisor o terceiro quartil, $5.500 \text{ kg}\cdot\text{ha}^{-1}$, que representaria uma produtividade maior e de maior notoriedade, seguindo o mesmo pressuposto de seleção de variáveis com 15% de significância, gerou um modelo com um índice de acerto de 91,5%, superior ao modelo com a mediana como divisor do banco de dados e altamente signicante, $<0,0001$ a qual explicou 74% da variabilidade dos dados de produtividade. Na Tabela 13 estão apresentados os resultados das variáveis agronômicas selecionadas pelo SAS, quanto maior o valor da relação canônica da variável agronômica maior a associação dela com a produtividade abaixo de $5.500 \text{ kg}\cdot\text{ha}^{-1}$ e quanto menor o valor da relação canônica da variável agronômica maior a associação dela com a produtividade acima de $5.500 \text{ kg}\cdot\text{ha}^{-1}$.

Tabela 13. Relação canônica com as variáveis agronômicas.

Variável	Canônica 1
pH CaCl2 10-20 cm	-0,429410
pH CaCl2 80-100 cm	-0,341630
P 0-10 cm	-0,417022
m% 0-10 cm	0,094520
S 80-100 cm	-0,470948
Cobre 10-20 cm	0,256390
(Cobre 10-20 cm) ³	0,199818
Resistência do solo 20cm	0,492691
Resistência do solo 50cm	0,434728
Temperatura média R1-R5	-0,085327
Temperatura mínima Semeadura-Emergência	0,0334
Déficit hídrico da semeadura a R8	0,296037

Regras de seleção:

Produtividade Abaixo 5.500 kg.ha⁻¹ = -562,69358*Constant + 26,39011*phcacl2_10-20cm + 83,31878*phcacl2_80-100cm + 0,08761*P_0-10cm + 14,37513* saturação_alumínio_0-10cm + 0,12323*Enxofre_80-100cm - 4,96762*Cobre_10-20cm + (0,01575*Cobre_10_20cm)³ - 2,93432*resistencia_solo_20cm + 13,67455*resistencia_solo_50cm + 32,31183*temperatura_média_R1-R5 - 9,23341*temperatura_mínima_Semeadura-Emergencia - 0,1366*déficit_hídrico_Semeadura-R8

Produtividade Acima 5.500 kg.ha⁻¹ = -606,48712*Constant + 29,26398* phcacl2_10-20cm + 86,93057* phcacl2_80-100cm + 0,10921*P_0-10cm + 15,34214*saturação_alumínio_0-10cm + 0,14804* Enxofre_80-100cm - 5,952* Cobre_10-20cm + (0,01904*Cobre_10_20cm)³ -6,62742* resistencia_solo_20cm + 11,19364* resistencia_solo_50cm + 33,67042* temperatura_média_R1-R5 - 9,94183*temperatura_mínima_Semeadura-Emergencia - 0,15659* déficit_hídrico_Semeadura-R8

Crítérios de seleção entre acima e abaixo da produtividade 5.500 kg.ha⁻¹:

Seleção = Produtividade Acima 5.500 kg.ha⁻¹ – Produtividade Abaixo 5.500 kg.ha⁻¹;

Se seleção < 0 = Produtividade Abaixo 5.500 kg.ha⁻¹

Se seleção > 0 = Produtividade Acima 5.500 kg.ha⁻¹

Na Figura 16 apresentam-se as variáveis canônicas 1 e 2 na classificação de acima e abaixo da produtividade de 5.500 kg.ha⁻¹ tendo o terceiro quartil como divisor dos valores, ou seja, 5.500 kg.ha⁻¹, e claramente a divisão é mais evidente.

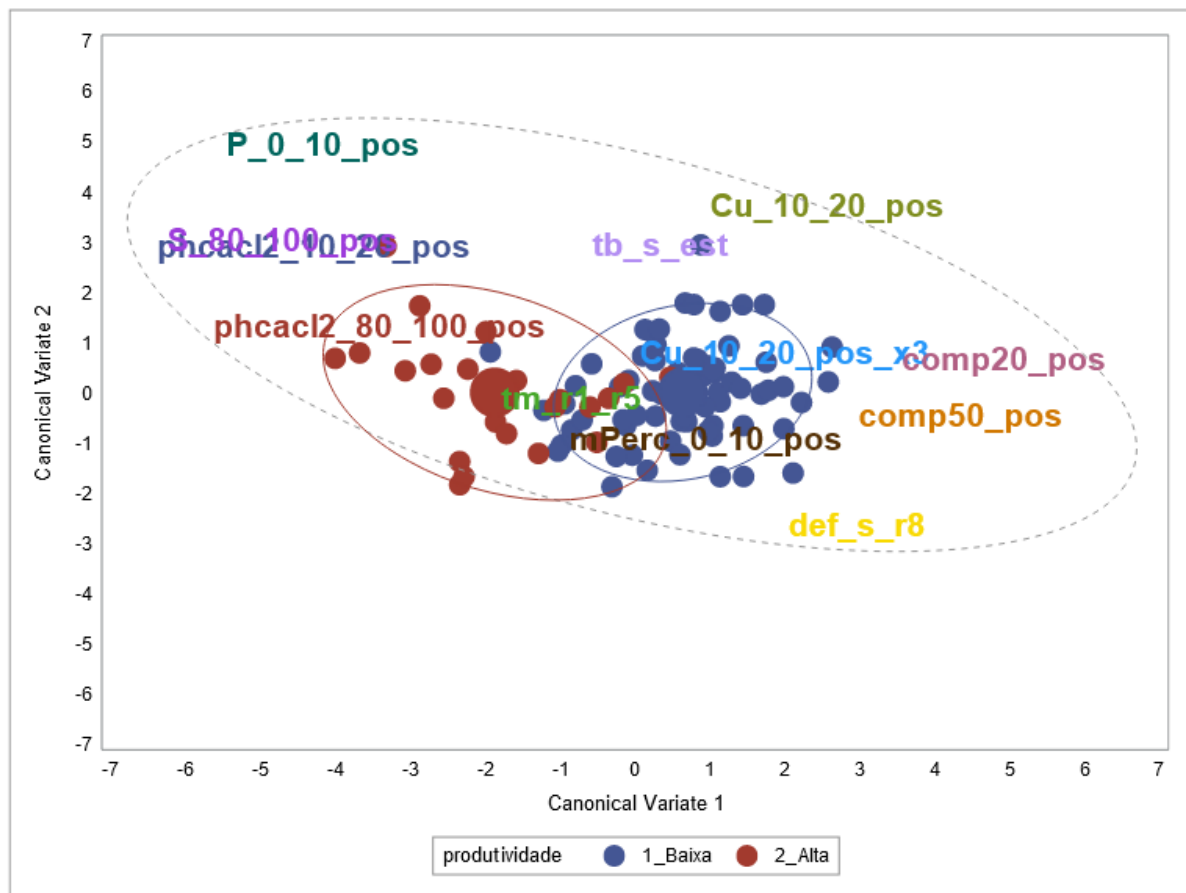


Figura 16. Variáveis canônicas 1 e 2 e a predição da produtividade em acima e abaixo de 5.500 kg.ha⁻¹.

As variáveis contempladas no modelo de 5.500 kg.ha⁻¹ possuem correlações agrônômicas com o estímulo ao crescimento radicular no perfil de solo, o que influencia no acesso à água e nutrientes pela lavoura. Dantas (2018), avaliando uma lavoura de 61 sc.ha⁻¹ em relação a 91 sc.ha⁻¹, identificou que 80% da evapotranspiração da lavoura com produtividade de 61 sc.ha⁻¹ extraiu água até a camada entre 10 e 20 cm de profundidade e na produtividade de 91 sc.ha⁻¹ obteve extração de água até a faixa entre 70 e 80 cm de profundidade. O crescimento radicular no perfil de solo e acesso a água e nutrientes possuem grande relevância para a alta produtividade.

Resistência do solo a 30 cm, 40 cm e 50 cm de profundidade: inúmeros trabalhos citam a relação dessa variável com o crescimento radicular e com produtividade. Rosolem e Takahashi (1998) avaliando 4 densidades diferentes de solo em condições controladas com soja, 1,06, 1,25, 1,43, 1,71 g.cm⁻³ identificaram a redução do crescimento radicular de 10 a 50% quando a resistência do solo aumentou de 0,52 MPa (1,45 g.cm⁻³) para 1,45 MPa (1,69 g.cm⁻³). Bengough e Young, (1993), avaliando diferentes níveis de resistência do solo, identificaram a redução de 55% no crescimento radicular quando a resistência do solo aumentou de 0,06 MPa para 1,8 MPa. Sako (2019) avaliou os valores de resistência de solo ao crescimento radicular de diversas lavouras de alta produtividade do Desafio de Produtividade, promovido pelo CESB, e todas as lavouras de 2014 até 2019 apresentavam valores de resistência de solo menores que 1,7 MPa. Busscher et al. (2000) identificaram que para cada 1 MPa reduzido houve um acréscimo de 720 kg.ha⁻¹. A resistência do solo, apesar de ser um indicador prático de levantar, tem a sua natureza variável, principalmente com a umidade, o que torna importante ponderar quando se faz esse levantamento, especialmente sobre a própria definição de um impedimento físico, que permeia por outros atributos para completar

a deficiência dessa metodologia, como porosidade, estrutura e dinâmica de expansão e contração da argila. No entanto, é notória sua relevância, uma vez que se identifica como um fator em comum nas produtividades acima de 5.500 kg.ha⁻¹ conforme apresentado nesse trabalho e outros citados (BUSSCHER; FREDERICK; BAUER, 2000; SAKO, 2019).

pH CaCl₂ a 10-20 cm e 60-80 cm; m% a 0-10 cm: A atividade dos nódulos é altamente depende da acidez do solo (BAKARI et al, 2020). De todos os nutrientes exportados da soja o nitrogênio é o sujeito ao maior volume e um dos mais desafiantes para o sistema de produção (BARTH et al., 2018; CIAMPITTI, SALVAGIOTTI, 2018). O alumínio, que é um dos componentes da acidez, é um grande detrator do crescimento radicular (NOBLE; SUMNER; ALVA, 1988), ele afeta o desenvolvimento dos citoesqueletos, processos metabólicos do cytosol e desregula a permeabilidade da membrana celular (SIVAGURU et al., 1999; ZHANG; RENGEL, 1999). Uma vez que o sistema radicular da soja chega a 3 m de profundidade é importante corrigir o alumínio no perfil do solo para permitir o crescimento e a atividade radicular (SAKO, 2019). A elevação do pH do solo gera os radicais OH⁻ e esses, por sua vez, reagem com Al²⁺ formando AlOH₂ que tem sua natureza precipitada (SANCHEZ, 2019). Sanzonowicz e Smith (1995), avaliando diferentes pH de 3,7 a 5,5 em subsuperfície no crescimento radicular de soja, constataram que para cada aumento de uma unidade de pH houve aumento linear de 26,6 cm por planta, enquanto a raiz secundária aumentou exponencialmente. Landell e colaboradores (2003) obtiveram correlação da produtividade de cana-de-açúcar no primeiro e segundo corte pelo pH H₂O na camada de 80-100 cm de profundidade. Battisti et al. (2016) realizaram um ensaio em Avaré, SP, com o objetivo de identificar a qual profundidade máxima de solo corrigido a soja responde em produtividade. Nesse ensaio, foi corrigido o solo de 20 em 20 centímetros até alcançar um metro de solo corrigido e, à medida que foram aprofundando a correção de solo, a produtividade aumentou, sendo o peso de mil grãos o componente evidente de alteração. Posteriormente, Sako (2019) instalou um ensaio em Itacemópolis, SP, fazendo a correção de solo de 50 em 50 centímetros, até 2 m de profundidade, e a soja continuou a responder em produtividade. Caires et al. (2008), avaliando a reaplicação de calcário sobre sistema de plantio direto com 3 Mg.ha⁻¹ em 1993 e 2000, mediram a acidez até 60 cm de profundidade e constataram a alteração dela, assim como a produtividade. Rheinheimer et al. (2000), avaliando um ensaio instalado em 1988 com 3,1 Mg.ha⁻¹ de calcário e sua reaplicação de 0; 3,6; 5,4 e 6 Mg.ha⁻¹ em uma única aplicação em 1994 e em 5 aplicações constataram uma frente alcalinizante provocando mudança na acidez até 60 cm de profundidade. A mobilidade do calcário no perfil do solo apresenta controvérsias, como no trabalho de Moraes e colaboradores (2007), avaliando a aplicação superficial de calcário com diferentes materiais vegetais em colunas de solo, buscando atingir 80% de saturação de bases, obtiveram alteração na acidez somente na camada superficial 0-8 cm de profundidade. Nota-se que nos estudos que avaliam a mobilidade de calcário são raras as avaliações da infiltração de água, porosidade contínua e estrutura do solo. Nesse estudo, dois fatores foram identificados na soja de alta produtividade, a resistência do solo e acidez em subsuperfície, merecendo essa reflexão. Na Figura 17 apresentamos uma avaliação de infiltração de solução com esmalte sintético branco numa soja de 50 sc.ha⁻¹ com restrição clara de infiltração de água devido à porosidade do solo e um solo de produtividade maior que 100 sc.ha⁻¹ com uma boa infiltração de água.

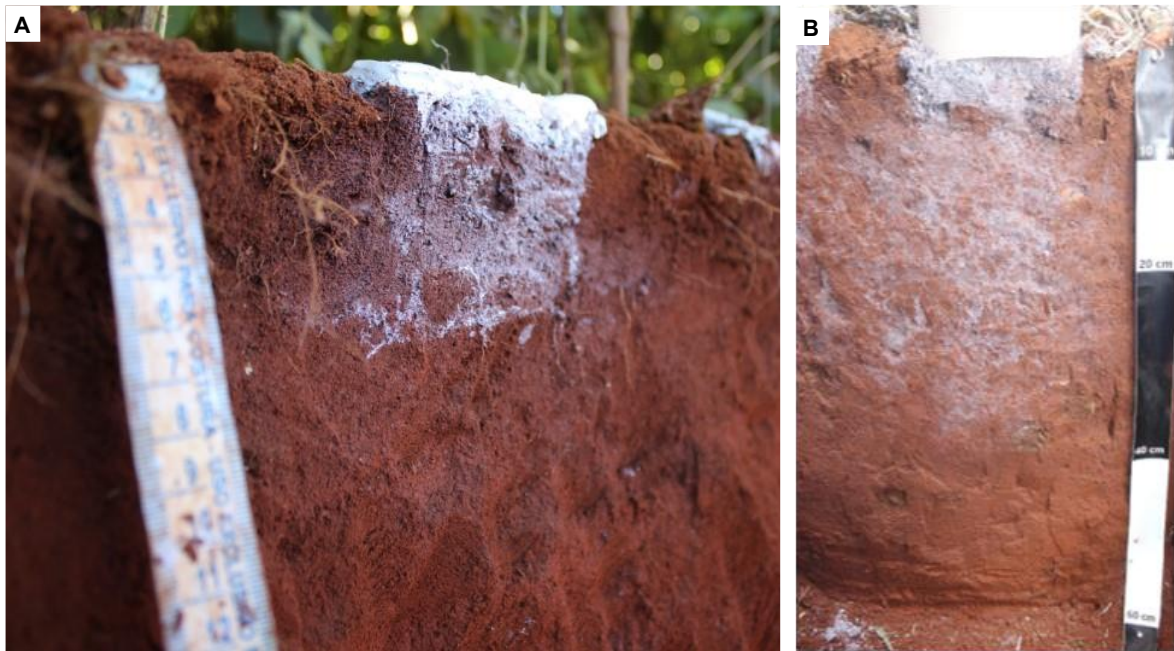


Figura 17. A) infiltração de uma solução com esmalte sintético branco numa lavoura de produtividade de 50 sc.ha⁻¹ (SAKO, 2019). **B)** infiltração de água numa lavoura com produtividade maior que 100 sc.ha⁻¹ (CESB, 2021).

Enxofre a 80-100 cm: o enxofre possui papel importante em carrear o alumínio tóxico a camadas mais profundas do solo, melhorando a tolerância ao alumínio. Guo et al. (2017), avaliando doses de enxofre em *Citrus grandis*, identificaram que a aplicação de enxofre reduziu os teores de alumínio nas folhas e nas raízes, aumentou o peso seco de raízes e folhas e reduziu a toxicidade de alumínio pela preservação do transporte de elétrons, no aparelho fotossintético por meio do metabolismo de enzimas e compostos antioxidantes e nas raízes na secreção de citrato. A presença de alumínio reduz a expressão da MicroRNA395, que suprime o transportador de sulfato 2, SULTR2;1, favorecendo a absorção de enxofre, com função importante na desintoxicação de alumínio no vacúolo das células radiculares (ALARCÓN-POBLETE et al., 2018). Uma vez que há maior dificuldade de corrigir a acidez em grandes profundidades do solo em relação às camadas superficiais, sugere-se que a importância dos teores de enxofre nessa camada de 80-100cm para a alta produtividade se deve justamente à sua relação com a tolerância ao alumínio e à redução dele nos tecidos foliares e radiculares (Figura 18).

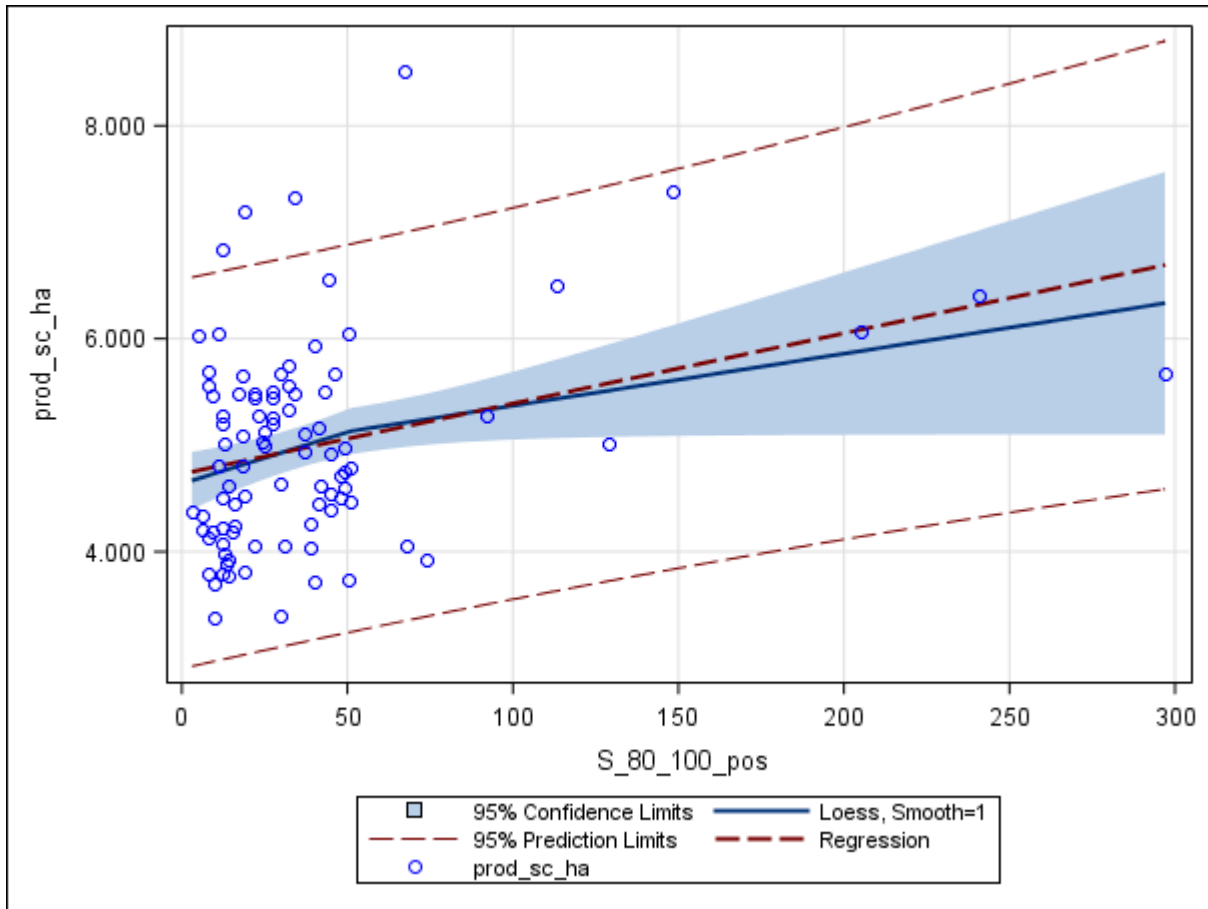


Figura 18. Teores de enxofre (mg.dm^{-3}) na camada de 80-100 cm de profundidade e produtividade da soja (kg.ha^{-1}).

Cobre a 10-20 cm: A amplitude dos teores de cobre no solo no trabalho levantado é grande indo de $0,1 \text{ mg.dm}^{-3}$ a $5,3 \text{ mg.dm}^{-3}$ e nessa faixa há teores considerados muito baixos e também teores muito altos elevando a toxidez. Nesse trabalho as produtividades acima de 4.920 kg.ha^{-1} estão com teores de cobre entre $1,3$ e $2,4 \text{ mg.dm}^{-3}$ dessa forma há tanto teores baixos quanto altos e por esse motivo a regressão ficou uma polinomial de 2º grau (Figura 19).

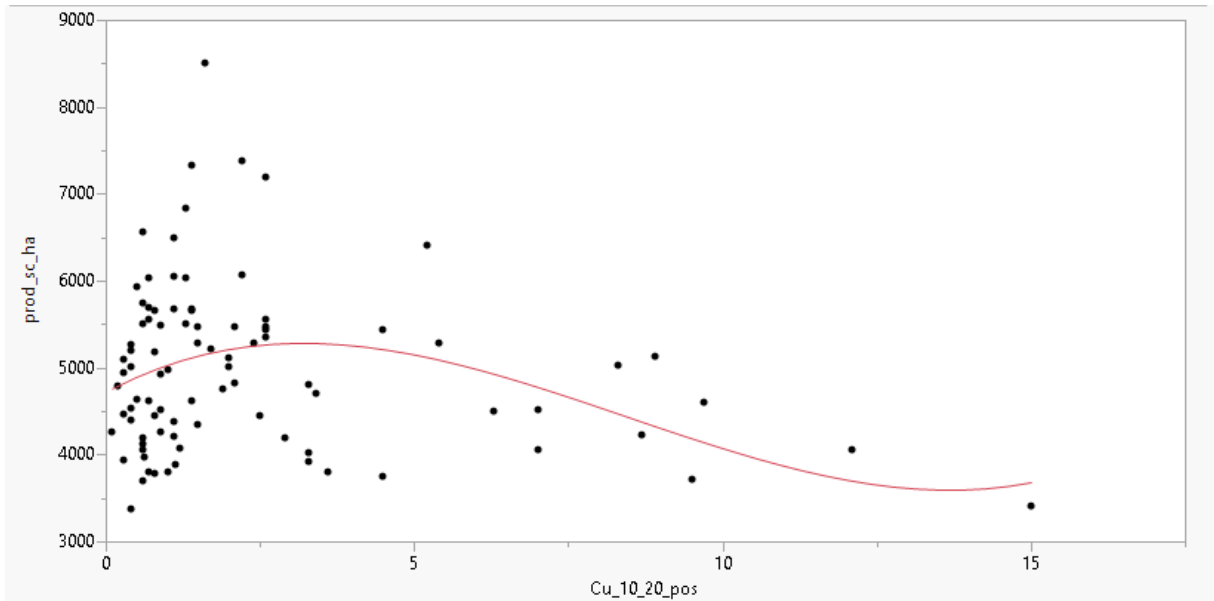


Figura 19. Teores de cobre (mg.dm^{-3} , DTPA) na camada de 10-20 cm de profundidade em relação à produtividade (kg.ha^{-1}).

Déficit hídrico da semeadura até R8: o déficit hídrico é uma métrica que reúne as variáveis de clima, profundidade efetiva das raízes e a capacidade de armazenamento do solo (BATTISTI, 2016). A presença de um bom volume de raízes no perfil do solo, somada ao baixo déficit hídrico, representam ganhos aditivos pelas duas variáveis em conjunto (Figura 20). Simpson e colaboradores (1979), avaliando a mistura de calcário em subsuperfície em relação à camada superficial do solo em alfafa, obtiveram de 100 a 200% de aumento de produção quando o solo se preservou úmido por todo o perfil, e quando se preservou úmido apenas na camada superficial, aumentou em 30% a produtividade. Uma das produtividades de maior notoriedade foi em 2014-15 em Ponta Grossa, PR, com 143 sc.ha^{-1} de soja: o perfil do solo dessa lavoura possui resistência de solo abaixo de 1,5 MPa, saturação de alumínio bem corrigido, bons teores de nutrientes, com o material genético NA5909 e baixo déficit hídrico (CESB, 2021).

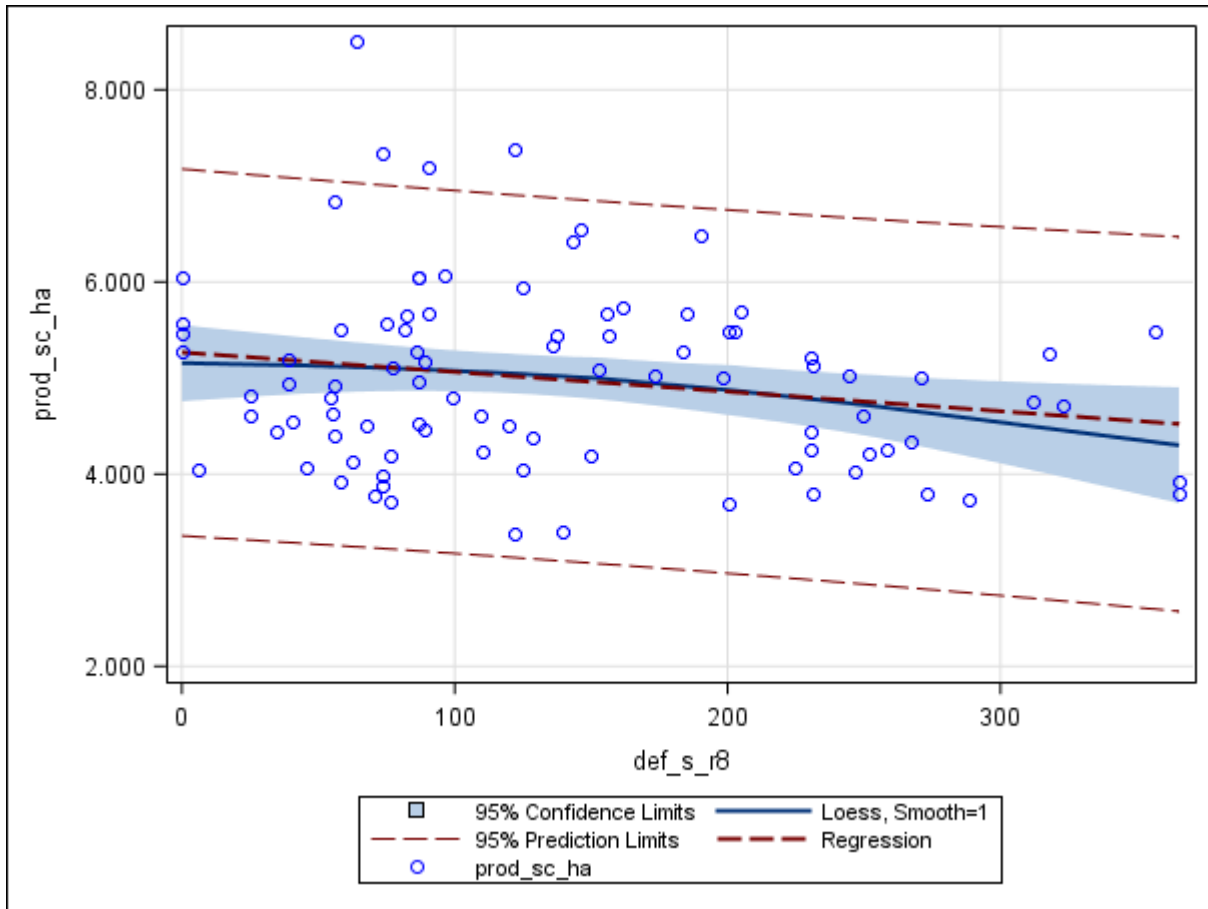


Figura 20. Déficit hídrico (mm de água) da semeadura até R8 e a penalização na produtividade da soja ($\text{kg}\cdot\text{ha}^{-1}$).

O presente trabalho possui uma natureza exploratória de excepcionais produtividades de soja na atual agricultura, sendo poucos os trabalhos que retratam esse patamar de produtividade (SAKO, 2019). Os 5 fatores agrônômicos citados são suportados por inúmeros trabalhos que mostram a relevância desses e a combinação desses fatores se destacou apresentada pelo modelo de Análise Discriminante através do bom índice de 91,5% de assertividade. Os teores de nutrientes no perfil do solo sugerem novas referências de fertilidade devido à mudança em todo o sistema de produção ao longo dos anos, principalmente a produtividade.

A “chave” para manejar uma lavoura e alcançar altas produtividades é justamente a habilidade do diagnóstico do ambiente de produção, solo – clima – planta, e nesse trabalho traz referências importantes para os técnicos terem sobre a fertilidade química e física do perfil do solo. O diagnóstico e a predição da produtividade pelo algoritmo permitirá o dimensionamento do investimento necessário para manejar o solo e o cálculo do retorno econômico.

5. CONCLUSÕES

i) O modelo de Análise Discriminante apresentou o melhor índice de assertividade (91,5%), tendo o valor de 5.500 kg.ha⁻¹ de soja como corte de classificação da produtividade da lavoura. A Análise Discriminante selecionou 7 fatores agronômicos: resistência do solo a 20 cm e 50 cm de profundidade; pH CaCl₂ na profundidade de 10-20 cm e 80-100 cm, saturação de alumínio a 0-10 cm, Fósforo 0-10 cm; Enxofre a 80-100 cm de profundidade; cobre a 10-20 cm de profundidade; e temperatura média do estágio fenológico de R1 a R5, a temperatura mínima do estabelecimento a emergência e o déficit hídrico durante o ciclo da soja.

ii) O levantamento da fertilidade do solo em superfície e principalmente em subsuperfície, a qualidade física do solo e os dados climáticos são de grande importância para compor modelos de boa previsão da produtividade de soja.

iii) O banco de dados não foi suficiente para aplicar os modelos de Floresta Randômica, Rede Neural e Support Vector Machine.

REFERÊNCIAS

- ALARCÓN-POBLETE, E. et al. **Molecular regulation of aluminum resistance and sulfur nutrition during root growth** *Planta* Springer Verlag, , 1 jan. 2018. Disponível em: <[https://research-repository.uwa.edu.au/en/publications/molecular-regulation-of-aluminum-resistance-and-sulfur-nutrition->](https://research-repository.uwa.edu.au/en/publications/molecular-regulation-of-aluminum-resistance-and-sulfur-nutrition-). Acesso em: 15 fev. 2021
- ALEKSANDER, I.; MORTON, H. **An introduction to neural computing**. [s.l.] Itp - Media, 1995.
- ALVES, G. R. **Estimativa da produtividade da soja com redes neurais artificiais**. [s.l.] Universidade Estadual de Goiás, 2016.
- ALVES, L. A. et al. **Quantificação da fixação biológica de nitrogênio para a soja em um Argissolo ácido sob semeadura direta após a calagem**. III ComSoja. *Anais...*Santa Maria, RS: 2020
- ARAÚJO, G. S. **Uso de Random Forests e Redes Biológicas na Associação de Polimorfismos à Doenças de Alzheimer**. [s.l.] Universidade Federal de Pernambuco, 2013.
- AVELAR, D. M. et al. **Resistência do solo em sistemas de manejo e rotação de culturas em Latossolo Vermelho**. X Jornada Acadêmica da Embrapa Soja. *Anais...*2014
- BARTH, G; FRANCISCO, E; SUYAMA, JT; GARCIA, F. Nutrient uptake illustrated for modern, high-yielding soybean. **Better Crops**. v. 102, n. 1, 2018.
- BATTISTI, R. **Calibration, uncertainties and use of soybean crop simulation models for evaluating strategies to mitigate the effects of climate change in Southern Brazil**. [s.l.] Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, 2016.
- BATTISTI, R. et al. **CIRCULAR TÉCNICA: INCREMENTO DE PRODUTIVIDADE DA SOJA: Redução do déficit hídrico associado à melhor distribuição e profundidade do sistema radicular**. [s.l: s.n.].
- BENGOUGH, A. G.; YOUNG, I. M. Root elongation of seedling peas through layered soil of different penetration resistances. **Plant and Soil**, v. 149, n. 1, p. 129–139, fev. 1993.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, out. 2001.
- BRIEDIS, C. et al. Can highly weathered soils under conservation agriculture be C saturated? **Catena**, v. 147, p. 638–649, 1 dez. 2016.
- BUSSCHER, W. J.; FREDERICK, J. R.; BAUER, P. J. Timing Effects of Deep Tillage on Penetration Resistance and Wheat and Soybean Yield. **Soil Science Society of America Journal**, v. 64, n. 3, p. 999–1003, 1 maio 2000.
- CAIRES, E. F. et al. Effects of soil acidity amelioration by surface liming on no-till corn, soybean, and wheat root growth and yield. **European Journal of Agronomy**, v. 28, n. 1, p. 57–64, 1 jan. 2008.
- CESB, C. E. S. B. **Fórum Nacional de Máxima Produtividade** Passo Fundo, RS, 2017.
- CESB, C. E. S. B. **Publicações - CESB – Comitê Estratégico Soja Brasil**. Disponível em: <[http://www.cesbrasil.org.br/publicacoes/>](http://www.cesbrasil.org.br/publicacoes/). Acesso em: 16 fev. 2021.

- CIAMPITTI, I. A.; SALVAGIOTTI, F. **New insights into soybean biological nitrogen fixation** *Agronomy Journal American Society of Agronomy*, , 1 jul. 2018. Disponível em: <<https://acess.onlinelibrary.wiley.com/doi/full/10.2134/agronj2017.06.0348>>. Acesso em: 15 fev. 2021
- CONAB, C. N. DE A. Acompanhamento Safra de Grãos - Safra 2017/18. **Acompanhamento da Safra Brasileira**, v. 5, n. 1, 2017.
- COOKE, G. W. **Fertilizing for Maximum Yield**. 3rd. ed. [s.l.] Collier Macmillan Canada, 1982.
- COX, M. S.; GERARD, P. D. Changes in yield classification in a soybean-rice rotation. **Precision Agriculture**, v. 11, n. 5, p. 507–519, 20 out. 2010.
- DA SILVA, I. N.; SPATTI, D.; FLAUZINO, R. **Redes neurais artificiais para engenharia e ciências aplicadas: curso prático**. [s.l.] Artliber Editora Ltda, 2010.
- DANTAS, J. P. DE S. **Perfil de extração de água do solo pela cultura de soja de alta e baixa produtividade de grãos**. [s.l.] Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, 2018.
- DE LIMA, R. P.; DE LEÓN, M. J.; DA SILVA, A. R. Comparação entre dois penetrômetros na avaliação da resistência mecânica do solo à penetração. **Revista Ceres**, v. 60, n. 4, p. 577–581, 2013.
- FURTADO, M. I. V. **Redes Neurais Artificiais: Uma Abordagem Para Sala de Aula**. Ponta Grossa, PR: Atena Editora, 2019.
- GERON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 1st. ed. Sebastopol, CA: O'Reilly Media, Inc., 2017.
- GUO, P. et al. Sulfur-Mediated-Alleviation of Aluminum-Toxicity in Citrus grandis Seedlings. **International Journal of Molecular Sciences**, v. 18, n. 12, p. 2570, 3 dez. 2017.
- HAMMER, R. G. **Modelagem da produtividade da cultura da cana de açúcar por meio do uso de técnicas de mineração de dados**. [s.l.] Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, 2016.
- JAYNES, D. B. et al. Cluster Analysis of Spatiotemporal Corn Yield Patterns in an Iowa Field. **Agronomy Journal**, v. 95, n. 3, p. 574–586, 1 maio 2003.
- KAUL, M. et al. Artificial neural networks for corn and soybean yield prediction. **Agricultural Systems**, v. 85, n. 1, p. 1–18, 2005.
- LANDELL, M. G. DE A. et al. Oxisol subsurface chemical attributes related to sugarcane productivity. **Scientia Agricola**, v. 60, n. 4, p. 741–745, 1 out. 2003.
- LEITE, J. G. D. B.; WAQUIL, P. D. Comportamento dos preços de produtos agrícolas: tendências, sazonalidades e choques. **Cadernos de Economia (Unochapecó. Impresso)**, v. 20, p. 97–117, 2006.
- LIAKOS, K. et al. Machine Learning in Agriculture: A Review. **Sensors**, v. 18, n. 8, p. 2674, 14 ago. 2018.
- MALAVOLTA, E. **Manual de nutrição mineral de plantas**. [s.l.] Editora Agronômica Ceres Ltda., 2006.
- MARSCHNER, P. **Mineral Nutrition of Higher Plants: Third Edition**. 3rd. ed. [s.l.] Elsevier Inc., 2011.
- MATSUMOTO, H. Cell biology of aluminum toxicity and tolerance in higher plants. **Int Rev Cytol**, v. 200, p. 1-47. 2000.

- MAZOYER, M.; ROUDART, L. **Histoire des agricultures du monde: du Neolithique à La crise contemporaine**. Paris: Seuil, 1997.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dez. 1943.
- MITCHELL, T. M. **Machine Learning**. [s.l.] McGraw Hill, 1997.
- MORAES, M. F. et al. Ionmobility in acid soils with surface application of lime, organic acid and crop residues. **Revista Brasileira de Ciencia do Solo**, v. 31, n. 4, p. 673–684, 2007.
- NARWAL, R. P.; KUMAR, V.; SINGH, J. P. Potassium and magnesium relationship in cowpea (*Vigna unguiculata* (L.) Walp.). **Plant and Soil**, v. 86, n. 1, p. 129–134, fev. 1985.
- NOBLE, A. D.; SUMNER, M. E.; ALVA, A. K. The pH Dependency of Aluminum Phytotoxicity Alleviation by Calcium Sulfate. **Soil Science Society of America Journal**, v. 52, n. 5, p. 1398–1402, 1 set. 1988.
- NYATSANGA, T.; PIERRE, W. H. Effect of Nitrogen Fixation by Legumes on Soil Acidity. **Agronomy Journal**, v. 65, n. 6, p. 936–940, 1 nov. 1973.
- OLIVEIRA, A. et al. **Marcha de absorção e acúmulo de macronutrientes em soja com tipo de crescimento indeterminado**. Reunião de pesquisa de soja. **Anais...**Londrina, PR: Embrapa Londrina, 2014
- RAIJ, B. van.; ANDRADE, J. C.; CANTARELLA, H.; QUAGGIO, J.A. Análise Química para Avaliação da Fertilidade de Solos Tropicais. Pesquisa Agropecuária Brasileira, ISBN 85-85564-05-9. 1999.
- REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-validation. In: LIU, L.; OZSU, M. T. (Eds.). **Encyclopedia of database systems**. 1st. ed. [s.l.] Springer Science+Business Media, LLC, 2009. p. 532–538.
- RHEINHEIMER, D. DOS S. et al. Aplicação superficial de calcário no sistema plantio direto consolidado em solo arenoso. **Ciência Rural**, v. 30, n. 2, p. 263–268, abr. 2000.
- ROBSON, A. D. (ED.). **Soil Acidity and Plant Growth**. [s.l.] Elsevier, 1989.
- ROSOLEM, C. A.; TAKAHASHI, M. Soil compaction and soybean root growth. In: **Root Demographics and Their Efficiencies in Sustainable Agriculture, Grasslands and Forest Ecosystems**. [s.l.] Springer Netherlands, 1998. p. 295–304.
- RUFFO, M.; HENNINGER, A.; BELOW, F. **A new experimental design to analyze the value of management factors contributing to high corn yield**. Crop Science Society of America International Annual Meeting. **Anais...**Long Beach, CA: 2010
- BAKARI, R; MUNGAI, N; THUITA, M; MASSO, C. Impact of soil acidity and liming on soybean (*Glycine max*) nodulation and nitrogen fixation in Kenyan soils, *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, v. 70:8, p. 667-678, 2020.
- RUSSELL, S.; PETER, N. **Artificial intelligence: a modern approach**. 3rd. ed. [s.l.] Pearson, 2010.
- SANCHEZ, P. **Soil Acidity. In Properties and Management of Soils in the Tropics**. Cambridge: Cambridge University Press, 2019. p.210-235. doi:10.1017/9781316809785.011

SAKO, H. et al. **Fatores decisivos para se obter produtividade de soja acima de 4.200 kg/ha. Circular Técnica CESB** (Comitê Estratégico Soja Brasil), , 2016.

SAKO, H. **Diagnóstico da qualidade física-biológica do solo e suas práticas de manejo**. Anais do Simpósio Caminhos para Soja de Alta Produtividade. **Anais...**Piracicaba, SP: 2019

SAKO, H.; SHIOZAKI, E. A.; DANTAS, J. P. S. Construção da fertilidade do perfil do solo para alta produtividade da soja. In: FANCELLI, A. L. (Ed.). . **Soja: condicionantes da alta produtividade**. Piracicaba, SP: [s.n.]. p. 53–68.

SANCHEZ, P. A. **Properties and Management of Soils in the Tropics**. [s.l.] Cambridge University Press, 2019.

SANTI, A. L. **Relações entre indicadores de qualidade do solo e a produtividade das culturas em áreas com agricultura de precisão**. [s.l.] Universidade Federal de Santa Maria, 2007.

SANZONOWICZ, C.; SMYTH, T. J. Effect of hydrogen on soybean root growth in a subsurface solution. **Pesquisa Agropecuária Brasileira**, v. 30, n. 2, p. 255–261, 1995.

SARTAIN, J.B., KAMPRATH, E.J. Effect of Soil Al Saturation on Nutrient Concentration of Soybean Tops, Roots, and Nodules1. **Agronomy Journal**, v. 69, p. 843-845. 1977.

<https://doi.org/10.2134/agronj1977.00021962006900050029x>

SAS Institute Inc., 2020, Cary, NC, USA.

SENTELHAS, P. C. et al. – Yield GAP – Conceitos, Definições e Exemplos. **Informações Agronômicas**, n. 155, p. 9–12, 2016.

SENTELHAS, P. C. et al. Clima e produtividade da soja: variabilidade climática como fator controlador da produtividade. **Boletim de Pesquisa. Fundação Mato Grosso.**, 2017.

SHIVAGURU, M., FUJIWARA, T., SAMAJ, J., BALUSKA, F., YANG, Z., OSAWA, H., MAEDA, T., MORI, T., VOLKMANN, D., MATSUMOTO, H. Aluminum-induced 1→3-β-D-glucan inhibits cell-to-cell trafficking of molecules through plasmodesmata. A new mechanism of aluminum toxicity in plants. **Plant Physiol**, v. 124, p. 991-1006. 2000.

SIMPSON, J. R.; PINKERTON, A.; LAZDOVSKIS, J. Interacting effects of subsoil acidity and water on the root behaviour and shoot growth of some genotypes of lucerne (*Medicago sativa* L.). **Australian Journal of Agricultural Research**, v. 30, n. 4, p. 609–619, 1979.

SIMONOVICOVA, M., TAMAS, L., HUTTOVA, J., MISTRÍK, I., 2004. Effect of aluminium on oxidative stress related enzymes activities in barley roots. **Biol Plant**, v. 48, p. 261-266. 2004.

SIVAGURU, M. et al. Impacts of aluminum on the cytoskeleton of the maize root apex. Short-term effects on the distal part of the transition zone. **Plant Physiology**, v. 119, n. 3, p. 1073–1082, 1999.

SMIDT, E. R. et al. Identifying Field Attributes that Predict Soybean Yield Using Random Forest Analysis. **Agronomy Journal**, v. 108, n. 2, p. 637–646, 1 mar. 2016.

TIMM, N. H. Discriminant and Classification Analysis. In: **Applied Multivariate Analysis. Springer Texts in Statistics**. [s.l.] Springer New York, 2002. p. 419–443.

VAN, G. rossum, & DRAKE, F. L. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.

VOLOUDAKIS, D. et al. Prediction of climate change impacts on cotton yields in Greece under eight climatic models using the AquaCrop crop simulation model and discriminant function analysis. **Agricultural Water Management**, v. 147, p. 116–128, 1 jan. 2015.

WU, H.; HUBBARD, K. G.; WILHITE, D. A. An agricultural drought risk-assessment model for corn and soybeans. **International Journal of Climatology**, v. 24, n. 6, p. 723–741, 1 maio 2004.

ZHANG, W. H.; RENGEL, Z. . Cytosolic Ca²⁺ Activities in Intact Wheat Root Apical Cells Subjected to Aluminium Toxicity. In: GISSEL-NIELSEN, G.; JENSEN, A. (Eds.). . **Plant Nutrition—Molecular Biology and Genetics**. [s.l.] Springer Netherlands, 1999. p. 353–358.