



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



MÓNICA ELIANA CÓRDOBA FIGUEROA

**TÉCNICAS DE AVALIAÇÃO NO ESTUDO DA QUALIDADE DE SEMENTES DE
TOMATE (*Solanum lycopersicum* L.)**

Botucatu

2024

MONICA ELIANA CORDOBA FIGUEROA

**TÉCNICAS DE AVALIAÇÃO NO ESTUDO DA QUALIDADE DE SEMENTES DE
TOMATE (*Solanum lycopersicum* L.)**

Tese apresentada à Faculdade de Ciências
Agronômicas da Unesp Campus de
Botucatu, para obtenção do título de Doutor
em Agronomia (Agricultura).

Orientadora: Prof. Dra. Maria Márcia
Pereira Sartori.

Botucatu

2024

F475t Figueroa, Mónica Eliana Córdoba
Técnicas de avaliação no estudo da qualidade de sementes
de tomate (*Solanum lycopersicum* L.) / Mónica Eliana Córdoba
Figueroa. -- Botucatu, 2024
74 p. : tabs., fotos

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrônômicas, Botucatu
Orientadora: Maria Márcia Pereira Sartori

1. Longevidade. 2. Tomate. 3. Tecnologia de sementes. 4.
Funções de ligação. 5. Qualidade fisiológica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências Agrônômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: TÉCNICAS DE AVALIAÇÃO NO ESTUDO DA QUALIDADE DE SEMENTES DE
TOMATE (*Solanum lycopersicum* L.)

AUTORA: MÔNICA ELIANA CÓRDOBA FIGUEROA

ORIENTADORA: MARIA MÁRCIA PEREIRA SARTORI

Aprovada como parte das exigências para obtenção do Título de Doutora em Agronomia (Agricultura),
pela Comissão Examinadora:

Pesquisadora Dr.ª MARIA MÁRCIA PEREIRA SARTORI (Participação Presencial)
Produção Vegetal / Faculdade de Ciências Agrômicas de Botucatu UNESP

Prof. Dr. ANDERSON RODRIGO DA SILVA (Participação Virtual)
Agronomia / Instituto Federal Goiano

Prof.ª Dr.ª RUTE QUELVIA DE FARIA (Participação Virtual)
Engenharia Agrícola / Instituto Federal de Educação, Ciência e Tecnologia Goiano

Prof. Dr. EDVALDO APARECIDO AMARAL DA SILVA (Participação Presencial)
Produção Vegetal / Faculdade de Ciências Agrômicas de Botucatu

Prof. Dr. ANTONIO ISMAEL INÁCIO CARDOSO (Participação Presencial)
Produção Vegetal / Faculdade de Ciências Agrômicas de Botucatu

Botucatu, 29 de novembro de 2023

Do meu anjo mais amado, minha irmã Deicy Tatiana Cordoba,

dedido

AGRADECIMENTOS

A Deus.

Aos meus pais Jesús Afranio Córdoba Santacruz e Alcira del Socorro Figueroa Davila, pela criação de duas filhas responsáveis, e pelo apoio incondicional na realização do meu sonho.

A minha irmã, Deicy Tatiana Cordoba Figueroa, pelo amor, pelo orgulho e por nunca soltar minha mão. Farei minhas conquistas as suas, por toda a minha vida, até o dia de nosso reencontro.

Ao Brasil, por abrir suas portas para que eu pudesse realizar o doutorado.

À Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, Faculdade de Ciências Agrônômicas – Campus de Botucatu.

Ao professor Amaral Da Silva, meu primeiro contato na UNESP, pela coorientação pelos ensinamentos em fisiologia sementes e por me fazer gostar ainda mais dessa área.

À professora Maria Márcia Pereira Sartori, pela excelente orientação, pelos ensinamentos em estatística (outra área que sou apaixonada), pelo apoio e por sempre acreditar em minhas capacidades.

À minha melhor amiga e irmã Vanessa Lopez Peñafiel, pela amizade incondicional, que, mesmo à distância, me soube oferecer apoio no desenvolvimento do doutorado.

Aos meus tios, primas e demais amigos que sempre questionaram sobre meus estudos e estavam na torcida por mim.

Às minhas melhores amigas brasileiras Brunna e Michelane, pelo amor compartilhado, pelo carinho, pelos ensinamentos. Vocês são minha família em Botucatu.

Ao grupo de pesquisa de Modelagem e Estatística Aplicada a Ciências Agrárias (MODELSTAT), especialmente ao Pedro Moreno quem cuida do meu experimento.

À Técnica do Laboratório de Sementes, Valeria Giandoni, pelo profissionalismo, amizade e ajuda na execução das análises.

A minha faculdade colombiana: Universidade de Nariño.

Ao Centro de Estudos Interdisciplinares Básicos e Aplicados - Fundação Ceiba e Gobernación de Nariño - Colombia, pelo empréstimo de estudos.

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

“Todos os dias sabemos mais e entendemos menos”

Albert Einstein

RESUMO

Na avaliação da qualidade fisiológica de sementes a longevidade, vigor, e germinação são de suma importância. A longevidade refere-se à capacidade de manter a viabilidade das sementes durante o armazenamento, e o tempo em que as amostras perdem 50% de viabilidade (P_{50}) ainda tem sido determinado por uma função de ligação Probit. No entanto, é observado, em vários casos, que o uso dessa técnica induz erro na predição do P_{50} . Estudos anteriores com outras funções foram aplicadas em sementes de soja e milho que exploraram alternativas robustas na predição desse parâmetro. Nas sementes de tomate, conhecidas pela germinação lenta e irregular, o condicionamento fisiológico “priming” é utilizado para acelerar e uniformizar o processo. A avaliação do efeito do “priming” demanda tempo e pessoal, assim novas abordagens da química analítica e métodos estatísticos podem diferenciar eficientemente sementes submetidas ou não ao “priming”. Portanto, os objetivos deste trabalho foram: (i) avaliar a função de ligação que melhor se ajusta no parâmetro de P_{50} e na longevidade de sementes de tomate em diferentes condições de armazenamento, para tanto, comparou-se as funções de ligação Logit, Cauchy e Cauchy-SSF com a usual Probit pelo método de ajuste de mínimos quadrados ordinários (OLS) e Logit e Cauchy com Probit pelo modelo linear generalizado (GLM) e; (ii) investigar o desempenho da técnica de espectrometria de emissão óptica com plasma induzido por laser (LIBS) combinada com análise multivariada de maneira rápida e com alta precisão na classificação de amostras de sementes primed e não primed. Os resultados demonstraram que (i) a função de ligação Cauchy-SSF, utilizando o método OLS, prediz com maior robustez a longevidade de sementes de tomate em diferentes condições de armazenamento em comparação com o Probit e, (ii) a técnica LIBS foi capaz de distinguir de forma precisa sementes primed e não primed de tomate com base na sua composição elementar, mostrando-se uma ferramenta de grande importância nessa avaliação.

Palavras-chaves: armazenamento; cauchy-ssf; composição elementar; *Solanum lycopersicum* L.; viabilidade; vigor.

ABSTRACT

In assessing the physiological quality of seeds, longevity, vigor and germination are of paramount importance. Longevity refers to the ability to maintain seed viability during storage, and the time at which samples lose 50% viability (P_{50}) has still been determined by a Probit link function. However, it has been observed in several cases that the use of this technique induces error in the prediction of P_{50} . Previous studies with other functions have been applied to soybean and corn seeds, which have explored robust alternatives for predicting this parameter. In tomato seeds, known for their slow and irregular germination, physiological priming is used to speed up and standardize the process. Evaluating the effect of priming takes time and personnel, so new approaches to analytical chemistry and statistical methods can efficiently differentiate between seeds that have been primed or not. Therefore, the objectives of this work were: (i) to evaluate the link function that best fits the P_{50} parameter and the longevity of tomato seeds under different storage conditions, to this end, the Logit, Cauchy and Cauchy-SSF link functions were compared with the usual Probit by the ordinary least squares (OLS) fitting method and Logit and Cauchy with Probit by the generalized linear model (GLM) and; (ii) to investigate the performance of the laser-induced breakdown spectrometry (LIBS) technique combined with multivariate analysis in classifying primed and non-primed seed samples quickly and with high precision. The results showed that (i) the Cauchy-SSF link function, using the OLS method, more robustly predicts the longevity of tomato seeds in different storage conditions compared to Probit and (ii) the LIBS technique was able to accurately distinguish primed and non-primed tomato seeds based on their elemental composition, proving to be a very important tool in this evaluation.

Keywords: storage; cauchy-ssf; elemental composition; *Solanum lycopersicum* L.; viability; vigor.

SÚMARIO

INTRODUÇÃO GERAL	17
CAPÍTULO 1 - FUNÇÕES DE LIGAÇÃO NA PREDIÇÃO DA LONGEVIDADE DE SEMENTES DE TOMATE.....	20
1.1 INTRODUÇÃO.....	22
1.2 MATERIAL E MÉTODO.....	24
1.3 RESULTADOS	29
1.4 DISCUSSÃO	37
1.5 CONCLUSÕES.....	40
REFERÊNCIAS	41
CHAPTER 2 - LASER INDUCED-BREAKDOWN SPECTROSCOPY (LIBS): AN ACCURATE TECHNIQUE TO SELECT PRIMED TOMATO SEEDS	43
2.1 INTRODUCTION.....	44
2.2 MATERIALS AND METHOD	45
2.3 RESULTS.....	49
2.4 DISCUSSION.....	56
REFERENCES	60
CONSIDERAÇÕES FINAIS	64
REFERÊNCIAS.....	65
APÊNDICE A – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 1 e 2.	66
APÊNDICE B – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 3 e 4.	67
APÊNDICE C – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 5 e 6.	68
APÊNDICE D – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 7 e 8.	69
APÊNDICE E – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 9 e 10.	70
APÊNDICE F – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 12.	71

APÊNDICE G – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 1, 2, 3 e 4.	72
APÊNDICE H – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 5, 6, 7 e 8.	73
APÊNDICE I – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 9, 10, 12.	74

INTRODUÇÃO GERAL

O tomate (*Solanum lycopersicum* L.) pertence à família das Solanáceas e é um vegetal considerado essencial em todo o mundo (Thwe *et al.*, 2020; Habibi *et al.*, 2023). Prevê-se que a produção e o consumo dessa cultura continuarão aumentando. Em países como o Brasil, sua produtividade média é de 65,14 t ha⁻¹, enquanto, na Colômbia é de 27,20 t ha⁻¹(Conab, 2019; Minagricultura, 2018).

As cultivares de tomate são destinadas a produção de frutos, com o objetivo de atender à demanda do mercado para comercialização “in natura” e/ou para a produção de molho e subprodutos na indústria. O custo da produção de sementes, usado na propagação do tomate, pode atingir valores elevados, devido ao nível tecnológico investido. Também pela redução da longevidade das sementes, o que se torna um grande problema no armazenamento (Batista, 2018; De Souza *et al.*, 2016).

A longevidade das sementes representa o período em que as sementes mantêm sua viabilidade e vigor após a maturidade fisiológica e é uma característica importante no âmbito ecológico, agrônômico e econômico (Hay *et al.*, 2019; Santos *et al.*, 2019). Durante o armazenamento a seco, as sementes perdem qualidade devido aos processos de deterioração, que são inevitáveis e irreversíveis (De Souza *et al.*, 2016). O primeiro indicador de deterioração é a queda do vigor, tornando-se mais sensível ao estresse durante a germinação, o que reduz a capacidade de germinar, e conseqüentemente, resulta na perda da germinabilidade em um período posterior de armazenamento (Bernal-Lugo e Carl Leopold, 1992; Bewley *et al.*, 2013; Rajjou; Debeaujon, 2008).

As propriedades das sementes, como composição lipídica, viabilidade inicial, estágio de maturação, massa de sementes, teor de óleo e composição de carboidratos afetam a longevidade das sementes e são fortemente influenciados pelas condições de umidade relativa e temperatura do ar durante o armazenamento. Sementes de diferentes espécies variam seu potencial de armazenamento; por exemplo, sementes de cebola e de soja possuem uma vida útil curta, enquanto sementes de cereais ou tomate podem ser armazenadas por mais tempo em condições adequadas.

A longevidade das sementes no armazenamento é uma característica quantitativa binária específica para cada espécie. Sua predição requer modelos matemáticos que expliquem o comportamento, bem como os fatores que podem influenciar a germinação ao longo do tempo de armazenamento. A função de ligação Probit determina o período em que 50% das sementes perderam viabilidade (P_{50}). Os dados além de binários nem sempre são normalmente distribuídos, e nesse caso não é possível utilizar a função Probit e, portanto, é necessário aplicar outras funções de ligação para prever a longevidade.

As funções de ligação Logit, utilizando o modelo linear generalizado, e Cauchy-SSF com mínimos quadrados ordinários, quando comparados com o usual Probit, fornecem resultados superiores na predição da longevidade de sementes de soja. Isso sugere a utilização dessas funções de ligação em outras espécies (Santos *et al.*, 2019) e representa uma inovação na utilização dessa metodologia para sementes de tomate.

Por outro lado, as sementes de tomate apresentam germinação lenta e irregular, resultando em problemas de propagação (Batista *et al.*, 2020). Para mitigar esse problema, empresas e pesquisadores usam técnicas de condicionamento fisiológico conhecido como priming. Esse tratamento é empregado em sementes de hortaliças e outras espécies que possuem sementes de tamanho pequeno, favorecendo a velocidade de germinação e emergência, resultando na uniformidade de germinação do lote e na formação de plântulas mais vigorosas.

Normalmente, o efeito de “priming” é determinado pela uniformidade da germinação e teste de vigor (t_{50}) das sementes, que demandam preparo de materiais, equipamentos, pessoal, tempo e minuciosos cuidados. No entanto, novas abordagens como a química analítica aliada a novas ferramentas de avaliação de dados como os métodos multivariados, podem permitir uma avaliação rápida e eficiente do efeito de “priming” pretendido, questão que ainda não foi estudada. Essa abordagem aqui foi investigada com a técnica de espectrometria de emissão óptica com plasma induzido por laser (laser-induced breakdown spectroscopy, LIBS), capaz de realizar a análise química direta de amostras sólidas e determinar simultaneamente vários elementos químicos em pouco tempo e com alta frequência analítica (Gamela *et al.*, 2019; Gamela *et al.*, 2020).

Frente ao exposto, os objetivos do trabalho foram: (i) avaliar e comparar as funções de ligação Logit, Cauchy e Cauchy-SSF com a usual Probit pelo método de mínimos quadrados – OLS e Logit e Cauchy com Probit pelo modelo linear generalizado – GLM na predição da longevidade de sementes de tomate em diferentes condições de armazenamento e; (ii) investigar o desempenho da LIBS combinada com análise multivariada de maneira rápida e com alta precisão na classificação de amostras de sementes com condicionamento fisiológico e não condicionadas em comparação aos métodos padrões. Além disso, identificar a composição elementar quando são submetidas ao “priming” e determinar conjuntamente com as variáveis fisiológicas sua relação nos dois grupos de amostras (com e sem “priming”).

CAPÍTULO 1

FUNÇÕES DE LIGAÇÃO NA PREDIÇÃO DA LONGEVIDADE DE SEMENTES DE TOMATE

Resumo

A longevidade de sementes representa a capacidade das sementes de permanecerem viáveis durante o armazenamento. A determinação da longevidade é usualmente realizada pela função de ligação Probit, que prediz o tempo em que 50% das sementes perdem sua viabilidade (P_{50}), e sua utilização tem predominado ao longo do tempo. No entanto, devido à natureza não normal do comportamento de sobrevivência das sementes, alguns pesquisadores não conseguiram realizar ajustes que proporcionassem resultados viáveis em seus dados, evidenciando a necessidade do estudo de outras funções para esta estimativa. Assim, foram realizados estudos sobre modificações a Probit ou a utilização de outras funções para predizer a longevidade. Nesse trabalho, avaliou-se as funções de ligação Logit, Cauchy e Cauchy-Santos-Sartori-Faria (SSF) pelo método de mínimos quadrados ordinários (OLS), bem como Logit e Cauchy pelo modelo linear generalizado (GLM), e comparou-se com Probit na predição da longevidade de sementes de tomate em diferentes condições de armazenamento. Quatro materiais vegetais foram utilizados e três soluções saturadas de cloreto de potássio (KCl), cloreto de sódio (NaCl) e cloreto de lítio (LiCl) a 35°C, que resultaram em condições de armazenamento com 80%, 75% e 60% de umidade relativa, respetivamente. O teor de água e a germinação foram avaliados no decorrer do armazenamento. Os dados foram transformados pelas funções de ligação Probit, Logit, Cauchy e Cauchy-SFF (OLS). Para a escolha da melhor função, usou-se as frequências dos valores de P_{50} dentro do Intervalo de Interesse Experimental (IIE), o Erro Médio Absoluto, Critério de Informação de Bayesiano e Deviance. Nove vezes, a função Cauchy-SSF determinou os valores de P_{50} dentro do IIE, superando a Probit. No modelo GLM as frequências dos valores de P_{50} dentro do IIE não apresentaram diferenças estatísticas significativas. Assim, a função Cauchy – SSF estimou a longevidade de sementes de tomate (P_{50}) em diferentes condições de armazenamento de forma robusta, quando comparada a Probit.

Palavras-chaves: P_{50} ; cauchy-ssf; probit; umidade relativa; armazenamento.

LINK FUNCTIONS IN PREDICTING THE LONGEVITY OF TOMATO SEEDS

Abstract

Seed longevity represents the ability of seeds to remain viable during storage. Longevity is usually determined using the Probit link function, which predicts the time at which 50% of the seeds lose their viability (P_{50}). However, due to the non-normal nature of seed survival behavior, some researchers have been unable to make adjustments that provide viable results in their data, highlighting the need to study other functions for this estimate. Thus, studies have been carried out on modifications to Probit or the use of other functions to predict longevity. In this work, the Logit, Cauchy and Cauchy-Santos-Sartori-Faria (SSF) link functions were evaluated using the ordinary least squares (OLS) method, as well as Logit and Cauchy using the generalized linear model (GLM), and compared with Probit in predicting the longevity of tomato seeds under different storage conditions. Four plant materials were used and three saturated solutions of potassium chloride (KCl), sodium chloride (NaCl) and lithium chloride (LiCl) at 35°C, which resulted in storage conditions with 80%, 75% and 60% relative humidity, respectively. Water content and germination were assessed during storage. The data was transformed using the Probit, Logit, Cauchy and Cauchy-SFF link functions (OLS). To choose the best function, we used the frequencies of P_{50} values within the Experimental Interval of Interest (EII), the Mean Absolute Error, Bayesian Information Criterion and Deviance. Nine times, the Cauchy-SSF function determined the P_{50} values within the EII, outperforming Probit. In the GLM model, the frequencies of P_{50} values within the EII showed no statistically significant differences. Thus, the Cauchy-SSF function estimated the longevity of tomato seeds (P_{50}) in different storage conditions in a robust way, when compared to Probit.

Keywords: P_{50} ; cauchy-ssf; probit; relative humidity; storage.

1.1 INTRODUÇÃO

A longevidade é definida como a capacidade de uma semente em permanecer viável por longos períodos durante o armazenamento a seco, sendo considerada uma característica importante na ecologia, agronomia e economia (Sano *et al.*, 2016; Santos *et al.*, 2019). Além disso, desempenha um papel fundamental na conservação de uma ou várias espécies que são relevantes ou apresentam características desejáveis para uma determinada região (Li; Pritchard, 2009).

A capacidade das sementes de permanecerem viáveis ao longo do tempo depende de vários fatores, como o teor de água, constituição química, estágio de maturação, viabilidade inicial, pressão de oxigênio, umidade relativa e temperatura de armazenamento (Walters, 1998; Groot *et al.*, 2012; Bewley *et al.*, 2013), sendo os dois últimos de maior relevância (Bewley *et al.*, 2013). Existem metodologias descritas que criam ambientes de umidades relativas e temperaturas em condições controladas de armazenamento, que permitem avaliar experimentalmente a perda da longevidade ao longo do tempo e interpretá-la corretamente, com a ajuda de funções de ligação.

Durante o armazenamento, a viabilidade das sementes diminui gradualmente devido a processos de envelhecimento ou eventos de deterioração. Esses eventos de deterioração consistem em um complexo de mudanças que ocorrem com o passar do tempo, causando diversos prejuízos aos sistemas e funções necessárias, resultando a morte celular e à perda de viabilidade da semente (Zinsmeister *et al.*, 2020). A intensidade e velocidade de deterioração são específicas para cada espécie e é irreversível, mesmo sob condições ideais de armazenamento.

A perda da viabilidade em sementes durante o armazenamento segue um padrão sigmoide e é possível determinar a curva de sobrevivência retirando amostras de sementes do armazenamento após diferentes períodos de tempo e avaliar sua germinação. Assim, sua predição comumente é realizada pela transformação do porcentual de viabilidade com o uso da função de ligação Probit, para linearizar os dados, a qual permite prever o tempo em que leva para a viabilidade diminuir para 50% (P_{50}) pelos coeficientes de regressão linear descrita na equação 1 de Ellis e Roberts (1980):

$$v = Ki - p \left(\frac{1}{\sigma} \right) \quad (1)$$

Onde, v representa o percentual de viabilidade em Probit, p é o período de armazenamento (dias), σ é o desvio padrão e Ki é o percentual de viabilidade em Probit no início do armazenamento.

O modelo Probit tem sido utilizado na predição do P_{50} de várias espécies (Santos, 2018). Sua vantagem está em levar em consideração a distribuição binomial do erro associado aos testes de germinação, o fato de que cada semente pode germinar ou não (Hay *et al.*, 2014). Além disso, pode modelar dados em que parece haver uma proporção constante de sementes presentes no lote que não fazem parte da população envelhecida de sementes, resultando em uma curva de sobrevivência assimétrica (Mead; Gray, 1999; Butler *et al.*, 2009; Hay *et al.*, 2014).

No entanto, o modelo Probit, ao ser ajustado por médias, pode levar a erros na predição dos valores do P_{50} devido os pressupostos exigidos pelo método ou pela regressão que são desconsiderados (os pressupostos de uma regressão linear permitem a utilização de médias ponderadas) (Amador *et al.*, 2011). Entretanto, ao desconsiderar as repetições, cria-se um efeito pela menor variação amostral, método usado por Ellis e Roberts (1980), o que, na obtenção do modelo ajustado, aumenta o coeficiente de determinação (r^2) e refere-se a um maior ajuste. Assim, correções podem ser realizadas na função Probit, assim como o uso de outras funções para uma estimativa mais robusta dos parâmetros estudados.

Desde 2018, têm-se estudado novas metodologias na predição do período de decréscimo de 50% da viabilidade (P_{50}) com funções simétricas de Logit e Cauchy-SFF, utilizando regressão linear e modelos lineares generalizados (Santos *et al.*, 2019), e assimétricas de Aranda-Ordaz, Complementar Log-Log, Logaritmica, Stukel, Weibull e Potência (Santos, 2018), utilizando o método de predição de regressão linear em sementes de soja.

Além disso, Lima (2021) aplicou as funções Logit e Cauchy-SSF na predição do P_{50} em ambientes controlados de armazenamento (80% UR e 40°C e 75% UR e 35°C) para sementes esverdeadas de soja. Lopes (2021) avaliou o P_{50} por meio de análise de

parâmetros bioquímicos de sementes de soja, bem como para a predição da germinação de diferentes acessos de mamona (T_{50}) (Siqueira, 2020). Essas pesquisas têm contribuído significativamente para o estudo da longevidade de sementes de soja e que podem ser aplicadas em outras espécies.

As metodologias anteriormente citadas não têm sido avaliadas em sementes de tomate. Nesse contexto, o objetivo desse trabalho foi avaliar e comparar as funções de ligação Logit, Cauchy e Cauchy-SSF com a usual Probit pelo método de mínimos quadrados – OLS e Logit e Cauchy comparada com Probit pelo modelo linear generalizado – GLM na predição da longevidade de sementes de tomate em diferentes condições de armazenamento.

1.2 MATERIAL E MÉTODO

1.2.1 Origem das sementes utilizadas

As sementes de tomate *S. lycopersicum* L. utilizadas neste trabalho foram dos acessos LA2285, LA1509 e LA1511, doadas pelo *Tomato Genetics Resource Center* (TGRC) e produzidas pelo grupo de pesquisa de sementes da Faculdade de Ciências Agronômicas (FCA) – Câmpus de Botucatu da Universidade Estadual Paulista (UNESP). A produção das sementes ocorreu em casa de vegetação na FCA, entre os meses de janeiro e maio de 2017 conforme descrito em Batista *et al.* (2020). Além desses acessos, foi usada a variedade Gaúcho obtida comercialmente. As sementes apresentavam forma ovalada com depressões laterais e área que variou de 3,18 a 6,80 mm². Assim, foi usado aproximadamente um total de 14.000 sementes e foram armazenadas a 12°C e 60% $\pm$ 2% de umidade relativa (UR) até o início do experimento.

1.2.2 Teste inicial

Teor de água (% base úmida) inicial das sementes: foi determinado em duas repetições de cinco sementes, utilizando o método da estufa a 105 $\pm$ 3°C por 24 horas segundo os critérios de ISTA (2020). O teor de água (% base úmida) inicial das sementes foi de 9,55%, 5,02%, 5,92% e 4,61% para LA2285, LA1509, LA1511 e Gaúcho, respectivamente.

Germinação: realizou-se quatro amostras de 25 sementes, que foram previamente desinfestadas com água e hipoclorito de sódio (0,5%) por 1 minuto, posteriormente distribuídas em caixas gerbox (11 x 11 x 3,5 cm) sobre duas folhas de papel filtro, umedecidas com água destilada equivalente a 2,5 vezes o seu peso seco (BRASIL 2009). Após isso, foram acondicionadas em câmaras tipo BOD (Biochemical Oxygen Demand), regulada com fotoperíodo de 8 h luz a 30°C e 16 h no escuro a 20°C de temperatura, optando-se como critério radícula $\geq 2\text{mm}$. A primeira contagem de germinação (G5) foi realizada aos 5 dias, a segunda (G14) aos 14 dias e a germinação total (TG) aos 20 dias. Para as análises, foi utilizada a soma da germinação das três contagens, conforme adaptado da Regra para Análise de Sementes (Brasil, 2009). Os resultados foram expressos em porcentagem

1.2.3 Estudo da longevidade

No estudo da longevidade, foram utilizadas três soluções: cloreto de potássio (KCl), cloreto de sódio (NaCl) e cloreto de lítio (LiCl). As duas primeiras foram preparadas mediante a adição gradativa do sal em 1 litro de água destilada até que o sal começasse a decantar no fundo do recipiente, tornando-se soluções saturadas (Bewley *et al.*, 2013). Foram usadas 800 g L⁻¹ de KCl e 1000 g L⁻¹ de NaCl. A solução LiCl foi preparada com aproximadamente 60 gramas do sal em 200 ml de água destilada.

Após a preparação das soluções, uma quantidade de 600 ml foi distribuída em caixas herméticas fechadas de plástico (10 x 10 x 5 cm) com suporte de tela e colocadas em condições controladas a uma temperatura de 35°C na câmara BOD, o que proporcionou dentro das caixas uma umidade relativa do ar de 80% para a solução de cloreto de potássio (KCl), 75 % para o cloreto de sódio (NaCl) e 60% para o cloreto de lítio (LiCl). Essas condições foram monitoradas com um datalogger da marca instrutemp (ITLO90).

Uma vez estabilizadas as umidades com a temperatura de 35°C dentre das caixas herméticas, 1200 sementes de cada material previamente distribuídas em sacos de tecido voil, foram suspensos sobre o suporte de tela (Figura 1) e armazenadas na BOD.

Figura 1 - Caixa hermética com solução saturada de KCl e suporte para os sacos com as sementes



Foto: Mônica Córdoba (2021)

No total, foram utilizados 12 tratamentos, resultantes dos ambientes proporcionados pelas três soluções saturadas a 35°C na BOD, juntamente com os quatro materiais vegetais listados na Tabela 1.

Tabela 1- Tratamentos resultantes das temperaturas e umidade relativa do ar do ambiente proporcionado pela solução salina e material vegetal utilizado

Tratamento	Solução Salina	Símbolo	Temperatura °C	Umidade Relativa do ar (%)	Material Vegetal
1	Cloreto de Potássio	KCl	35	80	LA2285
2	Cloreto de Potássio	KCl	35	80	LA1509
3	Cloreto de Potássio	KCl	35	80	LA1511
4	Cloreto de Potássio	KCl	35	80	Gaúcho
5	Cloreto de Sódio	NaCl	35	75	LA2285
6	Cloreto de Sódio	NaCl	35	75	LA1509
7	Cloreto de Sódio	NaCl	35	75	LA1511
8	Cloreto de Sódio	NaCl	35	75	Gaúcho
9	Cloreto de Lítio	LiCl	35	60	LA2285
10	Cloreto de Lítio	LiCl	35	60	LA1509
11	Cloreto de Lítio	LiCl	35	60	LA1511
12	Cloreto de Lítio	LiCl	35	60	Gaúcho

Os pontos amostrais foram realizados em intervalos de tempo aproximados de 5 a 10 dias, nos quais foram retiradas uma quantidade de 100 sementes, quatro repetições de 25 sementes de cada tratamento, para avaliar a germinação (conforme descrito anteriormente) até a mortalidade das sementes. O teor de água foi determinado em todos os pontos amostrais de longevidade, da mesma forma que foi descrito anteriormente.

Os dados de germinação obtidos ao longo do armazenamento foram expressos em porcentagem, sendo determinado os intervalos de P_{50} experimental, definido como intervalo amostral (em dias) em que as sementes perdem 50% da sua viabilidade durante o armazenamento.

1.2.4 Análise estatística

Modelos de regressão linear e modelos lineares generalizados foram ajustados aos dados de longevidade.

1.2.4.1 Métodos dos mínimos quadrados ordinários - OLS

Os dados ajustados pelo modelo de regressão linear passaram por transformações Probit e Logit. No caso do Probit, foi necessária a multiplicação dos dados para limitar a área da função entre $0 < x < 1$. Essa correção se faz necessária devido à propriedade inversa da normal ser infinita nos pontos 0 e 1. Assim, os valores (decimal da germinação) foram multiplicados por 0,9999999713, exceto nos valores de 0 e 1, que foram substituídos por $2,87 \times 10^{-07}$, o valor complementar a 1. Já para o Logit, foi realizado o mesmo processo, sendo que o valor para a multiplicação foi 0,9975 e o valor que substituiu os valores de 0 e 1 foi 0,0025. A função de Cauchy (equação 4), utilizada por Bonat *et al.* (2012) é descrita como a função trigonométrica tangente de um determinado produto (Tabela 2).

Após isso, os dados foram linearizados através das funções de ligação descritas na Tabela 2.

Tabela 2 - Funções de ligação no ajuste de modelos de regressão para dados de longevidade de sementes de tomate

Função de ligação	Modelo	Equação
Probit	$F(x) = \Phi^{-1}(x)$	(2)
Logit	$F(x) = \ln\left(\frac{x}{1-x}\right)$	(3)
Cauchy	$F(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$	(4)
Cauchy - SSF	$F(x) = \tanh\left(\pi\left(x - \frac{1}{2}\right)\right)$	(5)

A função Cauchy descreve a função trigonométrica tangente de um produto específico, que remete à círculos ou elipses gerando valores extremos nos pontos (0,1) inviabilizando seu uso (Santos, 2018). No entanto, Santos *et al.* (2019) propuseram uma correção possível para as caudas, substituíram a função trigonométrica pela função hiperbólica da tangente (equação 5), que remete a uma hipérbole e foi denominada Cauchy-Santos-Sartori-Faria (Cauchy-SSF).

1.2.4.2 Modelo linear generalizado – GLM

Foram utilizadas as funções de ligação Probit, Logit e Cauchy (Tabela 2), as quais são possíveis para resolver pelo modelo GLM.

Tanto no método OLS quanto no modelo GLM, foram calculadas a frequência em que o valor estimado de P_{50} ficou contido no IIE (período em dias em que se obteve experimentalmente os valores percentuais estimados desejados do 50% de germinação do lote). Essas frequências do usual Probit foram comparadas com as demais funções por meio do teste qui-quadrado, com nível de significância de 5%.

O Erro Médio Absoluto (EMA) também foi calculado para avaliar a precisão das funções na estimativa da longevidade das sementes de tomate (Equação 6):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Onde, n é o número total de observações, y_i é o valor observado e $\hat{y}_i$ é o valor previsto pelo modelo.

Além disso, foram considerados o Critério de Informação Bayesiano – BIC (Equação 7) e deviance (Equação 8).

$$BIC = n \ln(SQR/n) + n \ln(p) \quad (7)$$

Onde, SQR soma de quadrado do resíduo do modelo, p é o número de parâmetros do modelo, n é o tamanho da amostra.

$$Deviance = -2 (\ln L_{null} - \ln L_{model}) \quad (8)$$

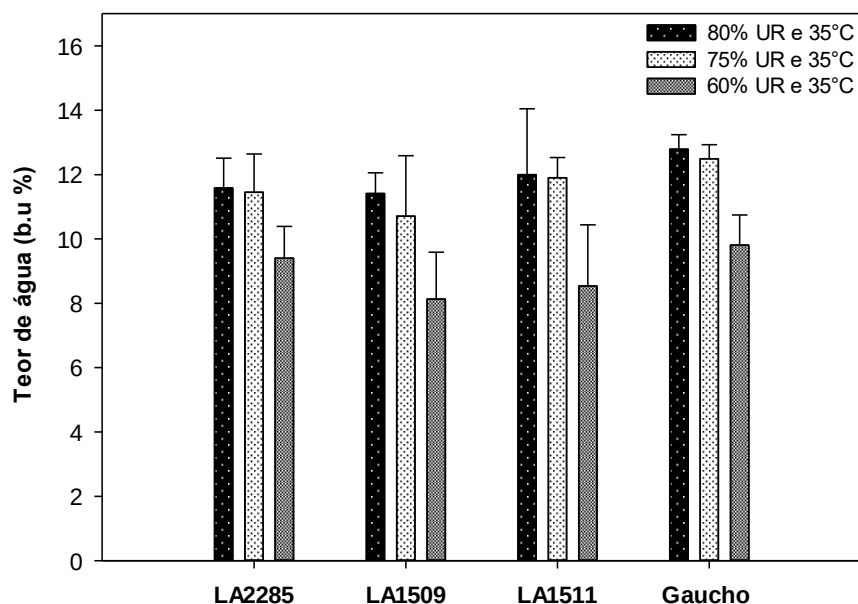
Onde, $\ln L_{null}$ é o logaritmo da verossimilhança do modelo nulo (sem preditores) e $\ln L_{model}$ é o logaritmo da verossimilhança do modelo ajustado.

Todas as análises dos dados foram realizadas em ambiente R versão 4.2.3.

1.3 RESULTADOS

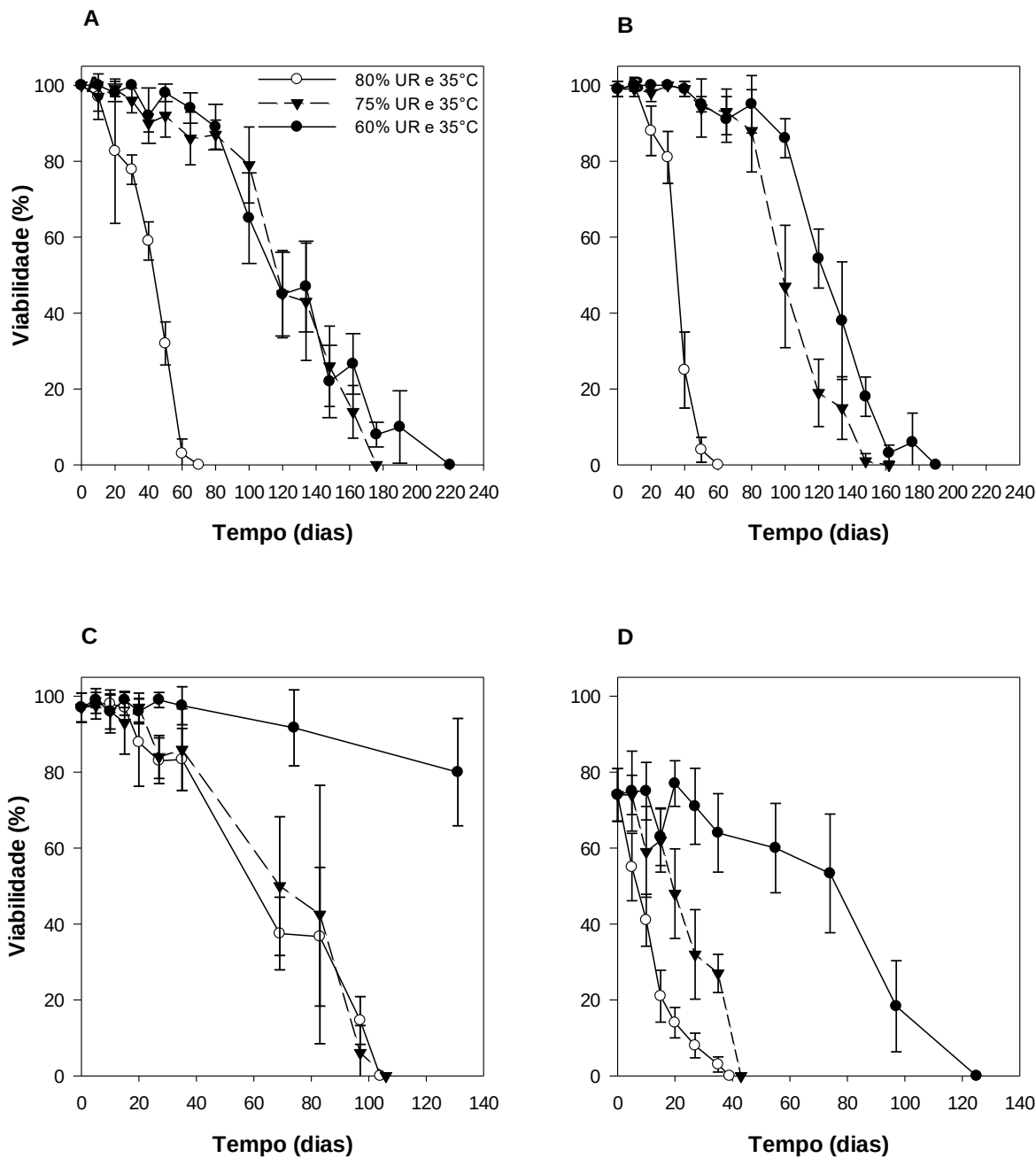
Após 27 dias de armazenamento controlado nos ambientes de 80%, 75% e 60% de UR e temperatura de 35°C, as sementes atingiram o teor de água de equilíbrio (% base úmida) conforme ilustrado na Figura 2.

Figura 2 - Teor de água (% base úmida) após 27 dias de armazenamento controlado (80%, 75% e 60% de UR e 35°C) em sementes de *S. lycopersicum*. As barras de erro mostram o desvio padrão



A germinação inicial para LA2285, LA1509, LA1511, e Gaúcho foi de 100%, 99%, 97% e 74%, respectivamente. No entanto, a capacidade germinativa ou viabilidade das sementes nas condições controladas de armazenamento foi diminuindo ao longo do tempo. Além disso, observaram-se diferentes comportamentos dos materiais vegetais dependendo das condições que se encontravam no armazenamento (Figura 3).

Figura 3 - Efeito do armazenamento controlado (80%, 75% e 60 % UR e 35°C) na viabilidade das sementes dos materiais vegetais LA2285 (A), LA1509 (B), LA1511 (C) e Gauchó (D) de *S. lycopersicum*. As barras de erro mostram o desvio padrão



As três condições controladas de armazenamento tiveram influência na viabilidade dos materiais vegetais estudados. Na condição controlada de 80% e 35°C as sementes perderam sua viabilidade em menos tempo. O material LA2285 levou cerca de 70 dias para a perda da sua viabilidade, o LA1509 em 60 dias, o LA1511 em 104 dias e, em menor tempo (39 dias) as sementes de Gaucho (Figura 3).

Em 75% e 35°C de armazenamento as sementes mantiveram sua viabilidade entre a faixa em dias de 106 e 176, exceto a Gaucho, que teve mortalidade das sementes aos 43 dias. Por outro lado, na condição de armazenamento de 60% e 35°C, fez com que o tempo de vida das sementes se prolongasse por 220 dias, 190 dias, 131 dias, respectivamente, para os materiais LA2285, LA1509, LA1511 e Gaucho (Figura 3).

Na Tabela 3 tem-se os parâmetros importantes de viabilidade inicial, intervalo de interesse (IIE) em dias em que P_{50} ocorreu, bem como a porcentagem associada a diminuição da viabilidade nesse intervalo, que permitiram verificar a predição do P_{50} .

Tabela 3 - Parâmetros de viabilidade inicial, intervalo de interesse experimental e percentual de protrusão do intervalo do IIE para cada tratamento

Tratamentos	Material Vegetal	Temperatura °C	Umidade Relativa do ar (%)	Viabilidade Inicial (%)	Intervalo de Interesse Experimental IIE (dias)	Percentual de Germinação do IIE (%)
1	LA2285	35	80	100	40-50	59-32
2	LA1509	35	80	99	30-40	81-25
3	LA1511	35	80	97	35-69	83-38
4	Gaucho	35	80	74	5-10	55-41
5	LA2285	35	75	100	100-120	79-45
6	LA1509	35	75	99	80-100	88-47
7	LA1511	35	75	97	35-69	86-50
8	Gaucho	35	75	74	15-20	62-48
9	LA2285	35	60	100	100-120	65-45
10	LA1509	35	60	99	120-134	54-38
11	LA1511	35	60	97	--	--
12	Gaucho	35	60	74	74-97	53-18

Os intervalos de interesse do P_{50} obtidos experimentalmente variaram entre os ambientes controlados de armazenamento. Nos tratamentos 1 a 4, o P_{50} esteve entre 5 e 69 dias. Por exemplo, no tratamento 3, o IIE foi de 35-69 dias, apresentando uma viabilidade inicial de 97%. Nos tratamentos 5, 6, 7 e 8, os IIEs do P_{50} variaram entre 15 e 120 dias, sendo que o tratamento 5 mostrou uma viabilidade maior em comparação com os outros, e também uma viabilidade inicial alta de 100%. Quanto aos tratamentos 9, 10 e 12, os IIEs de P_{50} foram de 100 a 120 dias, 120 a 134 dias e 74 a 97 dias, respectivamente (Tabela 3), sendo importante mencionar que o tratamento 11 foi desconsiderado das análises devido à falta de sementes, o que não permitiu ser avaliado até a perda total da viabilidade.

O K_i (coeficiente de viabilidade inicial) e $1/\sigma$ (desvio padrão da frequência de distribuição das sementes mortas) são parâmetros importantes na análise de regressão. Esses coeficientes, quando calculados em diferentes ambientes controlados de armazenamento usando o método OLS, tiveram valores com tendência semelhante para as funções Probit e Logit, porém, diferentes das Cauchy e Cauchy-SSF. No modelo GLM, os valores demonstraram uma tendência semelhante nas funções Probit, Logit e Cauchy. Dessa forma, a partir da sua relação, obtêm-se os respectivos P_{50} (Tabelas 4 e 7).

A função Cauchy-SSF prediz o maior número de valores de P_{50} dentro do intervalo de interesse experimental do que as funções Probit e Logit. Somente em dois momentos (tratamentos com condições controladas de armazenamento de 60% UR e 35°C) que não foi possível a predição utilizando essa função (Tabela 4).

Ao comparar os resultados preditos de P_{50} com o intervalo de interesse dos P_{50} experimentais (Tabela 4) observou-se que, um tratamento apresentou valor de P_{50} dentro desse intervalo usando a função Probit, o que correspondeu ao tratamento 4 (80 UR e 35°C). Três vezes foram estimados usando a função Logit, incluindo novamente o tratamento 4, bem como os tratamentos sete e oito (75% e 35°C). A Cauchy estimou cinco vezes, três na umidade de 75% e 35°C e dois de 80% e 35°C. E, finalmente a função Cauchy-SSF estimou um número maior de vezes (nove), representados por todos os tratamentos dos ambientes controlados de armazenamento 80 UR e 35°C e 75 UR e

35°C, exceto dois tratamentos de 60 UR e 35°C. Observou-se que os dois tratamentos (10 e 11) não foram estimadas de forma correta para nenhuma das funções.

Tabela 4 - Intervalo de interesse experimental (IIE) em dias, coeficientes de viabilidade inicial (Ki), o desvio padrão e tempo em que 50% das sementes perderam a viabilidade obtido pelo método mínimo ordinário - OLS pelas funções de ligação Probit, Logit, Cauchy e Cauchy-SSF

Trat	IIE (dias)	Probit			Logit			Cauchy			Cauchy-SSF		
		Ki	1/σ	P50	Ki	1/σ	P50	Ki	1/σ	P50	Ki	1/σ	P50
1	40-50	-1,28	-0,02	82	-0,95	-0,03	35	20658	590	35	1,22	-0,03	40
2	30-40	-2,06	-0,01	352	-1,83	-0,02	93	23531	-722	33	1,27	-0,04	34
3	35-69	-0,96	0,01	97	-0,30	0,00	195	10463	-145	72	1,12	-0,02	60
4	5-10	0,63	-0,09	7	0,94	-0,14	7	1244	-136	9	0,29	-0,04	7
5	100-120	-1,34	0,00	365	-0,83	0,00	969	16622	-165	101	1,26	-0,01	113
6	80-100	-2,61	0,00	991	-2,37	0,00	1091	24677	-260	95	1,31	-0,01	94
7	35-69	-0,89	-0,01	169	-0,53	-0,01	45	11328	-223	51	0,83	-0,01	64
8	15-20	1,61	-0,12	14	1,99	-0,14	15	5725	-487	12	0,62	-0,03	20
9	100-120	-2,53	0,01	363	-2,32	0,00	686	21890	-172	127	1,23	-0,01	117
10	120-134	-3,14	0,01	370	-3,02	0,01	567	25672	-222	116	1,31	-0,01	114
12	74-97	1,44	-0,04	34	1,93	-0,05	38	3679	-172	21	0,78	-0,01	62

Nota: Os valores de P₅₀ com negrito representam os valores preditos pela função dentro do IIE.

Segundo o teste qui-quadrado, a frequência do valor de P₅₀ contido dentro do IIE para a usual Probit apresentou diferença estatística significativa ($p < 0,05$) com os valores de P₅₀ preditos pela função Cauchy-SSF, no entanto, com Logit e Cauchy não diferiram entre si (Tabela 5).

Tabela 5 - Teste qui-quadrado (média) para as frequências de P₅₀ estimados pelas funções Probit, Logit, Cauchy e Cauchy-SSF dentro do intervalo de interesse

Função de ligação	Média das frequências
Probit	# 0,09
Logit	# 0,27
Cauchy	# 0,45
Cauchy-SSF	0,81

Nota: # não diferem entre si

O EMA, o BIC e Deviance para as funções de ligação na predição da longevidade das sementes de tomate nos diferentes ambientes controlados de armazenamento avaliados pelo método OLS, são apresentados na Tabela 6. Observou-se que, para todos os tratamentos, a função Cauchy – SSF apresentou melhores ajustes em comparação ao Probit, Logit e Cauchy.

Tabla 6 – Erro Médio Absoluto, Critério Bayesiano e Deviance das funções Probit, Logit, Cauchy e Cauchy – SSF pelo método OLS

Trat	MAE				BIC			Deviance			
	Probit	Logit	Cauchy	Cauchy-SSF	Probit	Logit	Cauchy-SSF	Probit	Logit	Cauchy	Cauchy-SSF
1	2,70	3,10	7569,59	0,22	168,52	178,90	17,14	262,20	362,70	2316809305	2,31
2	2,75	3,15	7568,89	0,29	149,22	158,76	30,54	236,69	332,71	2229503889	3,41
3	2,12	2,68	7656,76	0,17	199,86	218,71	-0,76	262,67	420,75	3450894613	1,74
4	0,34	0,41	2128,27	0,22	64,92	58,20	16,20	11,66	9,17	588779639	2,05
5	2,36	2,92	8204,23	0,25	285,07	307,81	36,42	429,35	644,43	4663871435	5,06
6	2,87	3,38	9004,90	0,26	274,64	294,49	40,43	476,77	698,37	5754740883	5,27
7	2,40	2,85	9844,73	0,43	206,19	221,04	75,33	307,67	445,95	4920068953	11,68
8	0,92	0,95	5200,01	0,16	119,44	120,35	11,35	43,15	44,25	1401880663	2,14
9	2,44	2,95	7914,18	0,22	324,60	348,83	23,56	491,70	718,02	5250301199	4,46
10	2,66	3,21	8143,94	0,26	311,35	334,33	44,76	513,31	752,88	5697208112	6,04
12	0,92	1,02	4476,54	0,20	151,19	159,47	17,75	61,83	74,63	1547509362	2,98

Nota: valores em negrito representam os melhores ajustes para cada modelo determinados pelos três critérios.

No modelo GLM, observou-se que, em nove dos tratamentos, o valor de P_{50} dentro do intervalo de interesse experimental foi estimado pela função de ligação Probit, enquanto para as funções Logit e Cauchy-SSF os valores estimados foram o mesmo número de vezes (oito) (Tabela 7). O teste qui-quadrado não apresentou diferenças estatisticamente significativas entre as funções de ligação nas frequências do valor de P_{50} dentro do intervalo de interesse. Novamente observou-se que o tratamento 12 não foi estimado de forma correta para nenhuma das funções utilizadas.

Tabela 7 - Intervalo de interesse experimental (IIE) em dias, coeficientes de viabilidade inicial (K_i), o desvio padrão ($1/\sigma$) e tempo em que 50% das sementes perderam a viabilidade obtido pelo método linear generalizado - GLM pelas funções de ligação Probit, Logit e Cauchy

Trat	IIE (dias)	Probit			Logit			Cauchy		
		K_i	$1/\sigma$	P50	K_i	$1/\sigma$	P50	K_i	$1/\sigma$	P50
1	40-50	2,70	-0,07	40	4,76	-0,12	40	6,53	-0,16	42
2	30-40	3,49	-0,10	34	6,55	-0,19	35	11,20	-0,31	36
3	35-69	2,03	-0,03	65	3,55	-0,05	65	5,23	-0,08	68
4	5-10	0,52	-0,08	7	0,92	-0,13	7	1,05	-0,16	7
5	100-120	2,60	-0,02	117	4,69	-0,04	119	6,88	-0,06	123
6	80-100	3,55	-0,04	100	6,74	-0,07	101	13,32	-0,13	100
7	35-69	1,19	-0,02	65	2,01	-0,03	65	2,41	-0,04	67
8	15-20	0,85	-0,05	17	1,40	-0,08	17	1,29	-0,07	17
9	100-120	2,83	-0,02	123	5,04	-0,04	124	6,79	-0,05	124
10	120-134	3,69	-0,03	121	6,98	-0,06	122	12,52	-0,10	125
12	74-97	0,89	-0,02	51	1,45	-0,03	51	1,34	-0,02	54

Nota: Os valores de P_{50} com negrito representam os valores preditos dentro do IIE.

A Tabela 8 apresenta o Erro Médio Absoluto, o Critério de Informação Bayesiano e Deviance dos modelos para os tratamentos avaliados pelo modelo GLM. Os três critérios de avaliação dos modelos indicaram que os melhores ajustes foram apresentados com Probit para os tratamentos 1, 3, 8, 9 e 12, com Logit para os tratamentos 2, 4, 5, 6 e 10 e com Cauchy para o tratamento 7.

Tabela 8 – Erro Médio Absoluto, Critério Bayesiano e Deviance das funções Probit, Logit e Cauchy pelo método GLM

Trat	MAE			BIC			Deviance		
	Probit	Logit	Cauchy	Probit	Logit	Cauchy	Probit	Logit	Cauchy
1	0,96	0,98	1,28	119,28	122,35	146,45	50,44	53,51	77,62
2	0,87	0,86	1,08	88,77	84,41	92,83	36,71	32,35	40,77
3	0,91	0,93	1,07	128,17	130,33	147,46	42,30	44,46	61,60
4	0,65	0,61	0,69	106,27	104,04	107,70	16,17	13,95	17,60
5	1,05	1,03	1,10	223,25	218,52	229,26	93,87	89,14	99,88
6	1,051	1,049	1,19	173,05	161,51	173,89	91,15	79,61	92,00
7	2,18	2,15	2,02	380,40	377,09	360,52	294,74	291,43	274,87
8	1,06	1,07	1,11	156,91	158,75	167,20	51,51	53,35	61,80
9	0,92	0,96	1,30	214,34	217,36	270,61	83,73	86,74	140,00
10	0,97	0,95	1,14	186,04	176,19	196,57	87,66	77,81	98,19
12	1,11	1,12	1,12	215,94	217,93	226,13	80,22	82,21	90,42

Nota: valores em negrito representam os melhores ajustes para cada modelo determinados pelos três critérios.

1.4 DISCUSSÃO

Dentre dos atributos de qualidade das sementes, encontra-se o atributo fisiológico, e nele a longevidade das sementes, representada pelo período de tempo em que uma semente se mantém viável durante o armazenamento. A viabilidade ao longo do armazenamento pode ser influenciada por vários fatores, inclusive se altera dependendo da espécie e dentro dela, ademais destacam-se principalmente a temperatura e umidade relativa como os mais importantes. Compreender e predizer a longevidade das sementes de tomate quando submetidas a ambientes controlados de armazenamento, não apenas utilizando o modelo comum que tem sido utilizado há anos (Probit), mas também como outros modelos ajudam na interpretação e compreensão mais precisa por ser comportamentos que ainda não foram explorados. Aqui, trazemos o entendimento obtido pelos métodos OLS (Probit, Logit, Cauchy e Cauchy-SSF) e o modelo GLM (Probit, Logit e Cauchy) na predição da longevidade de sementes de tomate em diferentes ambientes controlados de armazenamento.

Observou-se um comportamento inversamente proporcional entre a viabilidade das sementes de tomate no tempo e o ambiente controlado de armazenamento. Ou seja, as sementes armazenadas em condições controladas de 80% UR e temperatura de 35°C perderam sua viabilidade em menor tempo, enquanto à medida que a umidade diminuiu, a viabilidade dos materiais vegetais se prolongou por mais tempo (Figura 3A-D). Uma relação direta encontramos entre as três condições controladas de armazenamento e do teor de água das sementes, entre maior umidade relativa maior teor de água das sementes (Figura 2) e, conseqüentemente maior perda da sua viabilidade. Esse efeito é agravado pela alta temperatura, combinado com a alta umidade relativa resultando em períodos mais curto de armazenamento (Bewley *et al.*, 2013).

Importante ressaltar que, em condições de armazenamento com umidade relativa de 3 a 50% e temperatura de 5°C, as sementes de tomate possuem uma longevidade (P_{50}) em torno de 10 anos, conforme demonstrado por Fleming *et al.* (2019). No entanto, em nosso estudo ao serem armazenadas em condições controladas com diferentes umidades relativas (80%, 75% e 60%) e temperatura de 35°C, o P_{50} foi representado em

dias, diminuindo o tempo em que ocorreu o decréscimo de 50% da viabilidade dos materiais vegetais estudados.

Observamos que, ao reduzir a umidade relativa e usar uma temperatura de 35°C, o intervalo de interesse do P_{50} aumentou, ou seja, a viabilidade das sementes dos materiais vegetais se mantém viáveis por mais tempo (Tabela 4). Esse comportamento é consistente com o encontrado por Lima (2021) em sementes de soja, nas quais foram submetidas a dois ambientes controlados de armazenamento, a 75 % UR e 35°C, e 80 % UR e 40°C, onde o intervalo de interesse do P_{50} diminuiu com o aumento desses dois fatores.

Após as transformações dos dados de viabilidade com as funções de ligação usando o método OLS, a função Cauchy-SSF demonstrou maior precisão na predição dos valores de P_{50} dentro do intervalo de interesse experimental (Tabela 4). Isso pode ser atribuído a correção realizada por Santos (2018), que utilizou a função hiperbólica da função Cauchy, tornando-se Cauchy-SSF. Os autores observaram limitações do uso da função padrão (Cauchy), utilizada por Bonat *et al.* (2012), que mesmo depois da correção das caudas pela multiplicação de todos os pontos e substituição no ponto zero pelos valores de 0,95 e 0,05, verificou-se flutuações nos pontos 1 e 0. Ainda, Santos *et al.* (2019), avaliando a função Cauchy-SSF e comparando com Probit e Logit, encontraram que, em 10 lotes de sementes de soja, 20% dos lotes com valores de P_{50} dentro do II quando foi utilizado Probit, 40% para a função Logit e 80% para Cauchy-SSF. Isso indicou robustez na predição do P_{50} para sementes de soja utilizando o método OLS, o que corrobora com o encontrado nesse estudo com a predição da longevidade de sementes de tomate.

Além da predição com as maiores frequências do valor de P_{50} dentro do intervalo de interesse com o uso da função Cauchy-SSF, o Erro Médio Absoluto, o Critério Bayesiano e Deviance dos modelos permitiram determinar a robustez dessa função no estudo da predição da longevidade (P_{50}) das sementes de tomate (Tabela 6). O que indica uma acertada descoberta na utilização do método e da função de ligação na parte experimental ou em um banco de germoplasma de conservação de sementes de tomate.

O modelo GLM não permitiu identificar diferenças do número de vezes que o valor de P_{50} esteve dentro do intervalo de interesse ao usar os três modelos (Probit, Logit e Cauchy), ou seja, qualquer desses modelos pode ser usado para prever a longevidade das sementes de tomate nos ambientes controlados de armazenamento. No entanto, os critérios de avaliação para a escolha dos modelos foram inferiores a metade dos tratamentos para Probit e do Logit, e em apenas um tratamento para Cauchy (Tabela 8). Dessa forma, não houve um padrão claro de melhor desempenho de uma função de ligação em relação às demais.

No trabalho de Santos *et al.* (2019) com o método GLM e com os mesmos modelos em sementes de soja, houve um aumento no número de ocorrências do valor do P_{50} dentro do intervalo de interesse diferente com o encontrado nesse estudo. Isso possivelmente pode ser atribuído ao comportamento da longevidade em cada uma das espécies e os dados deste estudo não tiveram um bom ajuste nas três funções de ligação avaliadas.

É importante mencionar que o ambiente controlado de armazenamento de 60% UR e 35°C apresentou certa instabilidade em alguns momentos do armazenamento, o que prejudicou a tendência dos dados, causando ruído no momento da utilização das funções de ligação. Recomenda-se cautela em pesquisas futuras ao utilizar essa condição na avaliação da longevidade de sementes de outras espécies, incluindo o tomate. O oposto ocorreu com as outras duas condições (80% UR, 75% e 35°C) que apresentaram estabilidade. Isto possivelmente poderia ser em razão do número menor de ocorrências do P_{50} dentro do intervalo de confiança nessa condição de armazenamento (60% UR e 35°C) (Tabela 4 e 7).

1.5 CONCLUSÕES

Os materiais vegetais de tomate com diferentes contrastes de qualidade fisiológica (radícula $\geq 2\text{mm}$), incluindo a variedade Gaucho, utilizando o método de mínimos quadrados ordinários (OLS) com a função de ligação Cauchy – SSF prediz a longevidade de sementes de tomate com maior robustez em comparação com a função Probit.

O modelo linear generalizado (GLM) com as funções de ligação Probit, Logit, Cauchy pode ser usado para prever a longevidade de sementes de tomate, porém a escolha da função dependerá de qual função se ajusta melhor aos dados e dos critérios de avaliação que indiquem robustez na predição da longevidade de sementes de tomate.

REFERÊNCIAS

- AMADOR, J. P; LOPES, S.J; PEREIRA, J.E; SOUZA, A.M; TOEBE, M. Análise das premissas e adequação dos resíduos em modelo de regressão linear para valores individuais, ponderados e não ponderados utilizando SAS. **Ciência e Natura**, v. 33, p. 7–22, n. 2. 2011.
- BEWLEY, J. D; BRADFORD, K.J.; HILHORST, H.W; NONOGAKI, H. Seeds: physiology of development, germination and dormancy. 3 edition. New York. Springer Science & Business Media, 2013.
- KRZYŻANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. D. B.; MARCOS-FILHO, J. Vigor de sementes: conceitos e testes. Londrina. Copyright pela Associação Brasileira de Tecnologia de Sementes, 2020.
- BONAT, W.H; RIBEIRO JR, P. J; ZEVIANI, W.M. Regression models with responses on the unity interval: specification, estimation and comparison. **Biometric Brazilian Journal**, v. 30, n. 4, p. 415-431. 2012.
- BUTLER, L.H; HAY, F.R; ELLIS, R.H; SMITH, R.D; MURRAY, T.B. 2009. Priming and re-drying improve the survival of mature seeds of *Digitalis purpurea* during storage. **Annals of Botany**, v. 103, p. 1261–1270. 2009.
- ELLIS, R. H.; ROBERTS, E. H. Improved equations for the prediction of seed longevity. **Annals of Botany**, v. 45, n. 1, p. 13-30. 1980.
- FLEMING, M.B; HILL, L.M; WALTERS, C. The kinetics of ageing in dry-stored seeds: a comparison of viability loss and RNA degradation in unique legacy seed collections. **Annals of Botany**, v. 123, p. 1133–1146. 2019.
- GROOT, S.; SURKI, A.A., VOS, R. and KODDE, J. Seed storage at elevated partial pressure of oxygen, a fast method for analysing seed ageing under dry conditions. **Annals of Botany**. v. 110, p.1149–1159. 2012.
- HAY, F.R; MEAD, A; BLOOMBERG, M. Modelling seed germination in response to continuous variables: use and limitations of probit analysis and alternative approaches. **Seed Science Research**, v. 24, p. 165–186. 2014.
- LI, De-Zhu; PRITCHARD, H. W. The Science and economics of ex situ plant conservation. **Trends in Plant Science**, v. 14, n. 11, p. 614-621. 2009.
- LIMA, M. S. Influência do percentual de sementes esverdeadas de soja no comportamento da longevidade com diferentes ambientes de armazenamento. Dissertação (Mestra em Agronomia/Agricultura) Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, 2021.

LOPES, K.C. Predição da longevidade de sementes de soja por parâmetros bioquímicos. 2021. Dissertação (Mestra em Agronomia/Agricultura) Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2021.

MEAD, A and GRAY, D. Prediction of seed longevity: a modification of the shape of the Ellis and Roberts seed survival curves. **Seed Science Research**. V. 9, p. 63–73. 1999.

SANO, N. L. R; NORTH, H.M; DEBEAUJON, MARION-POLL, A, SEO, M. Molecular Aspects of Seed Longevity. **Plant and Cell Physiology**, v. 57, p. 660–674. 2016.

SANTOS, A.R.P. Funções de ligação na determinação da longevidade de sementes de soja. 2018. Dissertação (Mestra em Agronomia/Agricultura) Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2018.

SANTOS, A.R. P; FARIA, R. Q; AMORIM, D. J; GIANDONI, V.C. R; DA SILVA, E.A.A, & SARTORI, M.M.P. Cauchy, Cauchy–Santos–Sartori–Faria, logit, and probit functions for estimating seed longevity in soybean. **Agronomy Journal**, v. 111, n. 6, p. 2929-2939. 2019.

SIQUEIRA, O.A. Qualidade fisiológica e sanitária de acessos de mamoneira espontânea provenientes do Acre por modelos lineares generalizados. 2020. Dissertação (Mestre em Agronomia/Agricultura) Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2020.

WALTERS, Christina. Understanding the mechanisms and kinetics of seed aging. **Seed Science Research**, v. 8, n. 2, p. 223-244. 1998.

ZINSMEISTER, J; LEPRINCE, O; BUITINK, J. Molecular and environmental factors regulating seed longevity. **Biochemical Journal**, v. 477, n. 2, p. 305-323. 2020.

CHAPTER 2

LASER INDUCED-BREAKDOWN SPECTROSCOPY (LIBS): AN ACCURATE TECHNIQUE TO SELECT PRIMED TOMATO SEEDS

Abstract

The priming technique is widely used to accelerate and standardize seed germination. The priming effect is normally attested by determining germination uniformity and vigor test (t50). However, these tests require work and time, which suggests the need to develop fast and accurate methods to ensure the intended priming effect. In this study, an innovative technique capable of distinguishing primed and unprimed tomato (*Solanum lycopersicum* L.) seeds was analyzed based on their elemental composition, detected by the Laser-Induced Breakdown Spectroscopy (LIBS) combined with principal component analysis (PCA) and the classification models: linear, quadratic and multiple discriminant analyzes (LDA, QDA, MDA), and nearest neighbor methods (KNN). The seeds were subjected to priming and tests of germination and vigor were carried out. In total, 75 primed and 75 unprimed seed samples were used. The samples analyzed by LIBS mainly showed the chemical emission lines of the elements potassium, calcium and magnesium. PCA allowed the reduction of variables into two components, highlighting two distinct groups and explaining 94.4% of the total data variability. Precision, accuracy, Kappa, F1, sensitivity and specificity were > 90% in all models. The QDA model with a sensitivity and specificity of 100% can better classify primed and unprimed samples. Bitpot showed that the seeds subjected to priming have the intended effect. Thus, the LIBS technique combined with PCA and classification models demonstrated promising potential for distinguishing and selecting priming seeds.

Keywords: Classification models, Elemental composition, PCA, *Solanum lycopersicum* L., Vigor.

² Capítulo redigido de acordo com as normas do periódico **Seed Science Research**

2.1 Introduction

Tomato (*Solanum lycopersicum* L.) is among the main consumed vegetables on a global scale. It has great economic importance and is a source of several chemical elements and vitamins, in addition to being an anticancer agent (Mauro et al., 2020; Mahima et al., 2013).

In any crop, the physiological quality of the seeds is important for the establishment of seedlings and therefore their performance. Tomato seeds commonly have slow and irregular germination, resulting in propagation problems (Batista et al., 2020). To mitigate this problem, companies and researchers use physiological conditioning (priming) techniques. Among these techniques is the osmopriming method, which consists of controlling imbibition of the seeds in contact with aqueous solutions, allowing the water potentials to reach the pre-planned balance and activating the biochemical process in germination (phase I and phase II of acquisition of water in the seed), without primary root protrusion (phase III) (Marcos-Filho, 2015; Petronilio et al., 2021).

Physiological conditioning is mainly used in vegetable seeds and other small species, favoring the speed of germination and emergence with the formation of more vigorous seedlings (Batista et al., 2015). Furthermore, it has a positive effect on aged seeds, improving germination and producing strong seedlings with good physiological and biochemical performance (Habibi et al., 2023).

Normally, the priming effect is determined by the uniformity of germination and vigor test (t50) of the seeds, which require preparation of materials, equipment, trained professional, time and meticulous care. However, new approaches such as analytical chemistry combined with new evaluation tools, like multivariate methods, can allow a fast and efficient assessment of the intended priming effect.

One such approach includes the laser-induced breakdown spectroscopy (LIBS) technique, which consists of a laser source that, with the aid of focusing lenses, emits high power laser pulse focused onto the surface of the sample. These pulses cause fast heating, melting and vaporization of the sample, producing a plasma at high temperatures. The vaporized material (atoms, ions and molecules) is excited to a higher energy level and when returning to a lower energy level (spontaneous emission), it emits specific

wavelength of radiation for each element (Noll and Noll, 2012; Pasquini et al., 2007; Castro, 2016).

The main benefits of LIBS are direct analysis of solid samples without preparation, small sample quantity, high analytical frequency, multi-element character, obtaining precise and accurate data and reduced risk of loss of analytes and contamination (Costa et al., 2019; Machado et al., 2020). However, LIBS presents limitations in the interactions between the laser and the sample matrix, which can hinder the specific analysis of the analyte. Pre-processing with the use of normalizations, multivariate analysis with principal component analysis and classification methods with linear, quadratic and multiple discriminant analyzes (LDA, QDA and MDA), and nearest neighbor methods (KNN) are widely used in the development of analytical methods by LIBS, thus improving its performance (Pořízka et al. 2018; Larios et al. 2020).

Due to the countless benefits of LIBS, the scientific community has shown interest in this technique and it has become a tool applied in agriculture, proposing innovations and improvements. In the field of seeds, only one paper was found, where it was investigated the performance of LIBS to differentiate high and low vigor soybean seed lots (Larios et al., 2020). Thus, the present study aims to investigate the performance of LIBS combined with multivariate analysis, in a fast and accurate way, to classify seed samples with physiological conditioning and unconditioned, in comparison to standard methods. Furthermore, it seeks to identify the elemental composition when subjected to priming and determine, together with the physiological variables, their relationship in two groups of samples (with and without priming).

2.2 Materials and Method

The *S. lycopersicum* seed samples used in this study were accessions LA0358, LA2285, LA1509, LA1511 and LA1620, donated by the Tomato Genetics Resource Center (TGRC) and produced by the seed research group at the Faculty of Agricultural Sciences – Botucatu Campus from the São Paulo State University (UNESP), as described in Batista et al. (2020). The seeds had an oval shape with lateral depressions and an area

range from 3.18 to 5.28 mm². A total of 2,500 seeds were used, which were stored at 12°C and 60% ± 2% relative humidity (RH) until the start of the experiment.

Priming

Out of 1,400 tomato seeds, 750 were osmoprimed according to Batista et al. (2020), which consisted of placing the seeds in tubes containing 15 mL of polyethylene glycol 600 (PEG 6000) solution with an osmotic potential of -1.0 MPa at 20°C for 60 h in the dark. To avoid the lack of oxygen in the solution during incubation, small orifices were made in the tube caps. Subsequently, the tubes were placed on a mixer (Multifunctional mixer MR-II model Biomixer) to stir the solution during incubation time. The PEG solution was renewed three times to avoid changes in the osmotic potential. After osmopriming, the seeds were washed with distilled water, placed on paper towels for 24 h at 20°C and 60% ± 2% relative humidity (RH). The water content (% wet basis) was 15.61% for the primed seeds, and 12.58% for the unprimed ones.

The amount of PEG 6000 was estimated from research published by Villela et al. (1991):
$$OP = (-1.18 \times 10^{-2}) \times C - (1.18 \times 10^{-4}) \times C + (2.67 \times 10^{-4}) \times C + (8.39 \times 10^{-7}) \times C^2T \quad (1)$$

Where OP means osmotic pressure (measured in bar), temperature is indicated by T in °C and PEG concentration is represented by C in g. kg⁻¹ of water. 284.021g of PEG 6000 were used.

Physiological Analysis

Seed germination and vigor

The primed and non-primed seeds were characterized by the germination test, choosing the primary root length ≥ 2mm as a criterion. Thirteen replications of 25 seeds from each treatment were distributed in gerboxes (11 x 11 x 3.5 cm) on two sheets of filter paper, moistened with distilled water equivalent to 2.5 times of their dry weight (BRASIL, 2009). Afterwards, they were placed in a germination chamber, regulated with a photoperiod of 8 h in the light at 30°C and 16 h in the dark at 20°C. The first germination

count (G5) was performed on day 05, and the total germination (TG) on day 14. The results were expressed as a percentage.

The vigor test was performed with the same number of repetitions and in the same method as previously described. After 24 hours, seed germination was evaluated at 6-hour intervals until reaching 50% (t50). The calculation was carried out using cumulative germination data, with the Germinator curve fitting module (Joosen et al., 2010).

LIBS analysis

Samples and LIBS

As amount of 75 prepared and 75 unprepared seed samples were directly verified by the LIBS instrument, J200 model from Applied Spectra (Fremont, USA). The operating conditions were: Nd:YAG laser with a wavelength of 1064 nm; *spot size* of 50 μm ; *delay time* of 1.9 μs and laser pulse energy of 53 mJ. With the described parameters it was possible to obtain a theoretical fluence around 1000 J cm^{-2} . In total, 750 spectra were acquired (5 spectra per sample) using the raster analysis method. For this instrument, the emission was collected from range 186 to 1042 nm. The detection system consists of 6 spectrometers: (i) 186 to 311 nm (0.059 nm resolution); (ii) 311 to 465 nm (0.073 nm resolution); (iii) 465 to 591 nm (0.062 nm resolution); (iv) 591 to 693 nm (0.050 nm resolution); (v) 693 to 884 nm (0.094 nm resolution) and (vi) 884 to 1042 nm (0.079 nm resolution). A total of 12,288 variables were recorded.

Statistical analysis of physiological experiments

The normality and homogeneity of the data were verified using the Shapiro-Wilk and Bartlett tests with 5% significance. The data met the assumptions of normality and homogeneity. Therefore, primed and unprimed seed vigor data (t50) were compared using the t test and a 5% level of significance. For the germination data, the generalized linear model (GLM) was fitted and compared using the Chisq test and a 5% level of significance.

Statistical analysis of data obtained by LIBS

A total of 750 spectra made it possible to identify the intensity of the emission lines of the chemical elements from the primed and unprimed seeds, which allowed to build a set of data to evaluate their differences and their discrimination. The data were normalized using the arithmetic mean, providing a matrix of 150 samples by 12,288 variables (150:12288) with the help of Matlab software (R2019b, MathWorks, USA) and Microsoft Excel for data organization. Subsequently, PCA was applied to reduce the dimensionality of the data and construct linear combinations. That is, the orthogonal vectors called principal components (PCs) or dimensions that are ordered by the amount of explanation of the data variance, and are not correlated with each other (Mingoti 2005; Abdi et al., 2013).

The number of components was decided based on the screeplot graph (graph of eigenvalues of a matrix), when the variance explanation percentage is stabilized and the curve is flattened, representing the number of components that have a significant effect on the new data (Abdi et al., 2013). The PCA approaches to data analysis used in this study were successfully studied in a recent paper by Ferreira et al., 2022.

In the PCs, the coordinate values of the observations are called scores, which are new data for each sample in each component. These values were used to build the classification model. The scores for each sample allow us to interpret how each variable or group of variables influenced the variance of the data, that is, how each emission spectrum of the chemical elements of the seeds is influencing the classification of the groups and which are the most important for the representativeness of the data.

The classification models were built based on the scores of the new variables. The total dataset was randomly divided into training (70%) and testing (30%) data. The training data was used in the calibration and in the building of each model, and afterwards, the test data was used to evaluate the performance and classification metrics and, therefore, it was not considered in models building. In such wise, 106 samples were randomly selected to compose the training dataset, and 44 samples for the testing dataset.

Four classification models were analyzed: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), mixture discriminant analysis (MDA) and K-nearest neighbors (KNN). Model parameters were defined based on cross-validations ($n=2$). The

performance of the classification models was evaluated based on the parameters: precision, accuracy, Kappa, F1, sensitivity and specificity (Hossin & Sulaiman, 2015), which allowed to determine the best model for classifying primed and unprimed seeds. The confusion matrices of the results of each model were evaluated. On the other hand, the elements present in the primed and unprimed seed samples were identified based on the highest emission peaks. Thus, with the areas of these elements and the physiological variables that had a statistically significant difference, they were used to build a new PCA-Biplot, with the aim of analyzing the relationships between samples and variables and verifying the influence of each variable. These analyzes were carried out in the software R version 4.2.3., using the package “FactoMineR” and “Caret”.

2.3 Results

Priming effect

The osmopriming of -1.0 MPa at 20°C of the PEG solution did not show a significant effect ($p \leq 0.05$) on the total germination of tomato seeds (Fig. 1A). The opposite happened in the first count, where there was a higher germination rate (83%) at a percentage of 15% in relation to the non-primed seeds (Fig. 1B). Primed seeds started germination after 54h of soaking, and reached 50% of germination (t_{50}) faster (144h), while non-primed seeds started germination after 96h of soaking and took another 24h to reach the same percentage of germination (Fig. 2A). There was also a significant reduction in t_{50} ($p \leq 0.05$) after 29h in relation to unprimed seeds (Fig. 2B), that is, priming accelerated the germination process.

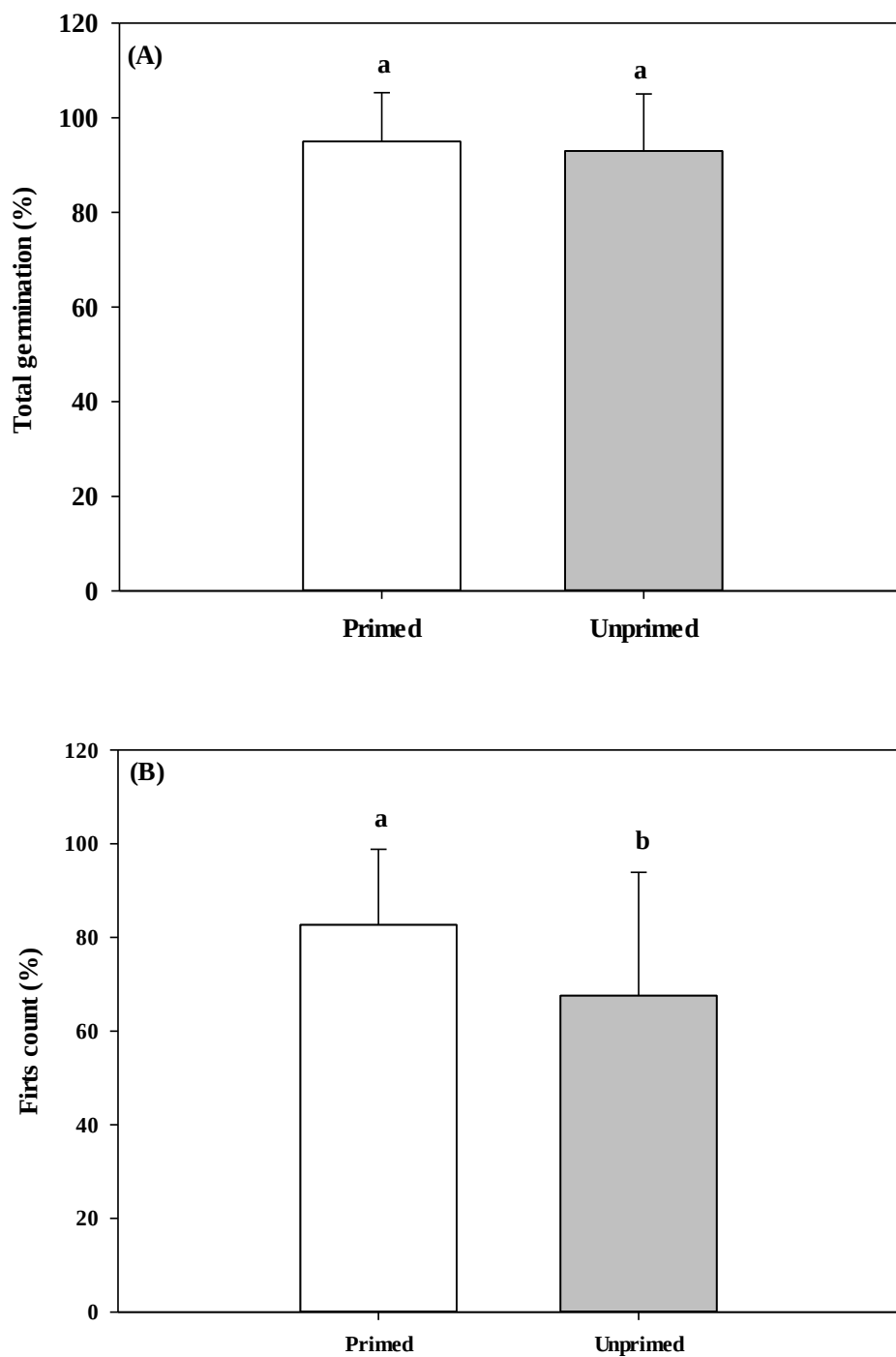


Fig. 1. Effect of priming on *S. lycopersicum* seeds. Total seed germination (A) and first count (B) on primed and unprimed seeds. Different letters show significant difference ($p \leq 0.05$) by Chisq test ($n=13$). Error bars show standard deviation.

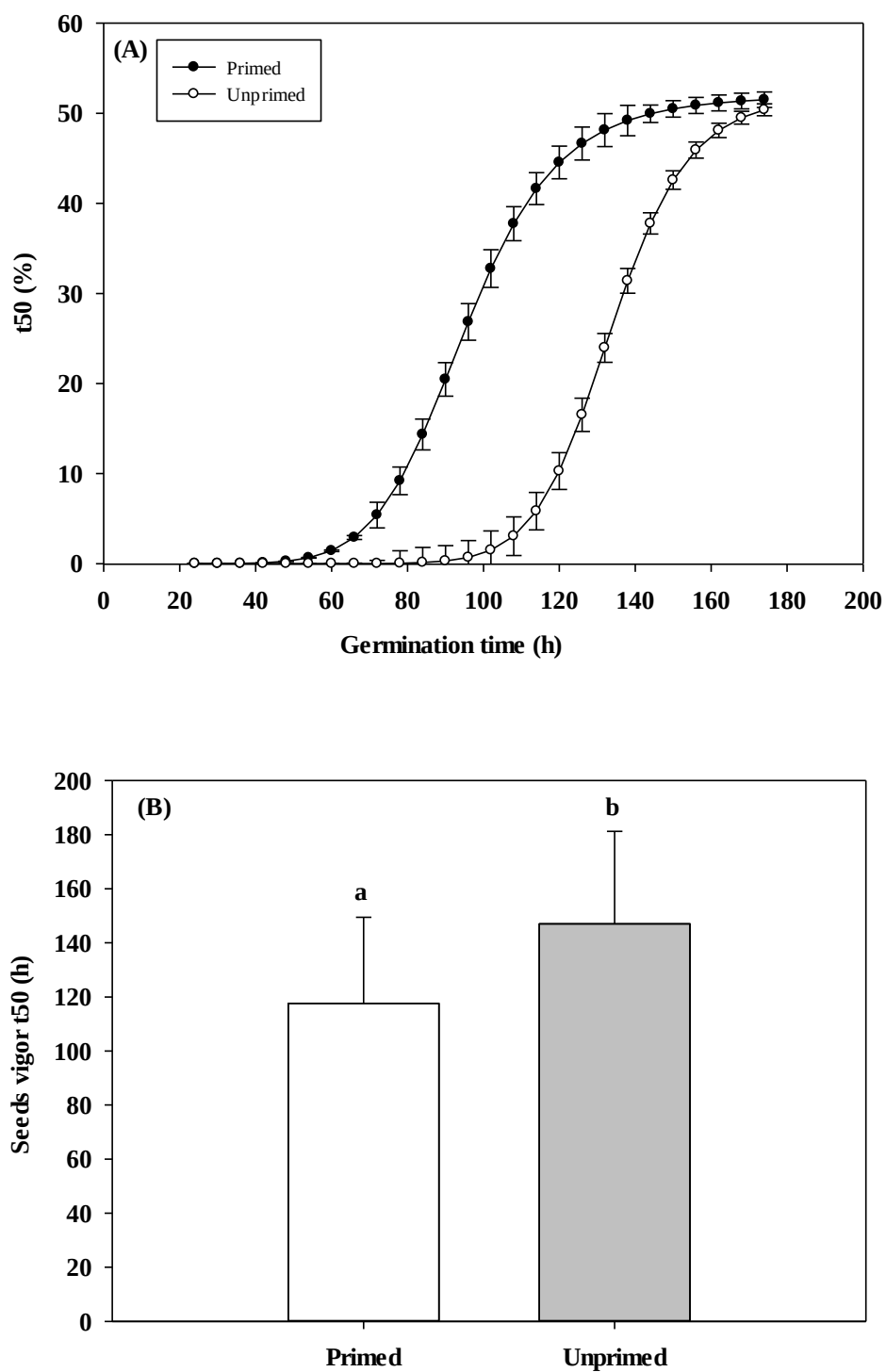


Fig. 2. Effect of priming on *S. lycopersicum* seeds. Curve of 50% germinated seeds (A) and seed vigor (B) in primed and unprimed seeds. Different letters show significant difference ($p \leq 0.05$) by t test ($n=13$). Error bars show standard deviation.

LIBS Element emission lines

The LIBS spectra acquired in the 75 primed and in the 75 unprimed seed samples showed the chemical emission lines of the elements, that is, the seeds are mainly composed of potassium (K), calcium (Ca), magnesium (Mg) and sodium (Na) (Fig.3). However, the Na emission signal showed interference due to sample handling, therefore this element was disregarded.

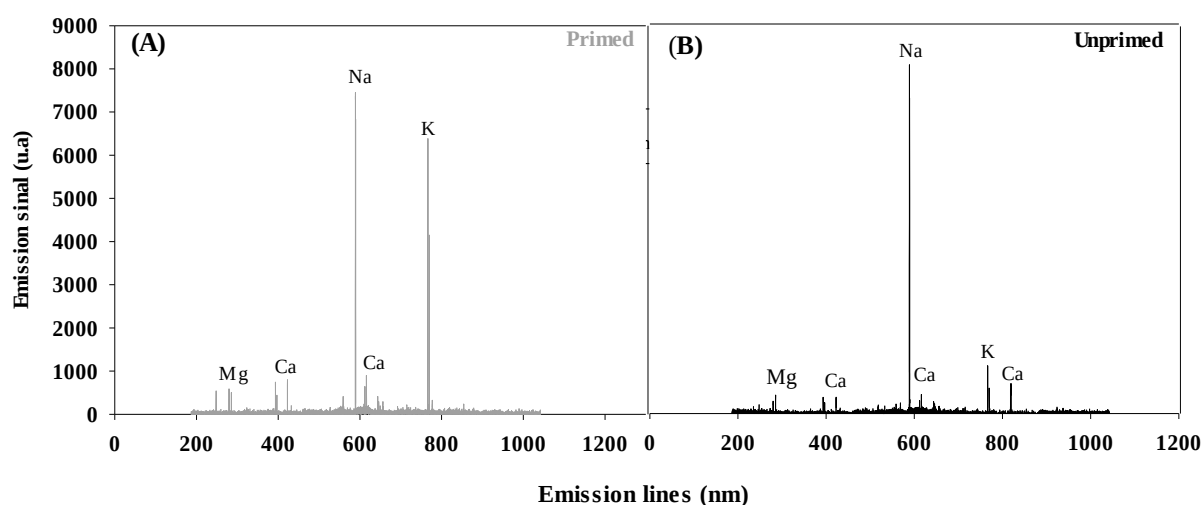


Fig. 3. LIBS emission spectrum of a primed (A) and unprimed (B) seed sample.

Principal Component Analyses (PCA)

The application of PCA performed satisfactorily in reducing the dataset, reducing 12,288 variables to just two principal components, with 94.4% of the total explanation of data variability. The first component represented 92.2% of the explanation contribution, and the second one represented 2.2% of the variance.

The two PCs have scores that revealed a clear tendency of separation and distinction between the two groups of primed and unprimed seeds (Fig. 4). These scores made it possible to make inferences about the behavior of the samples. In this case, the primed samples (seeds) present a uniform and larger grouping compared to the non-primed group, which presented a greater dispersion in the graph, resulting in a lower uniformity of the group itself.

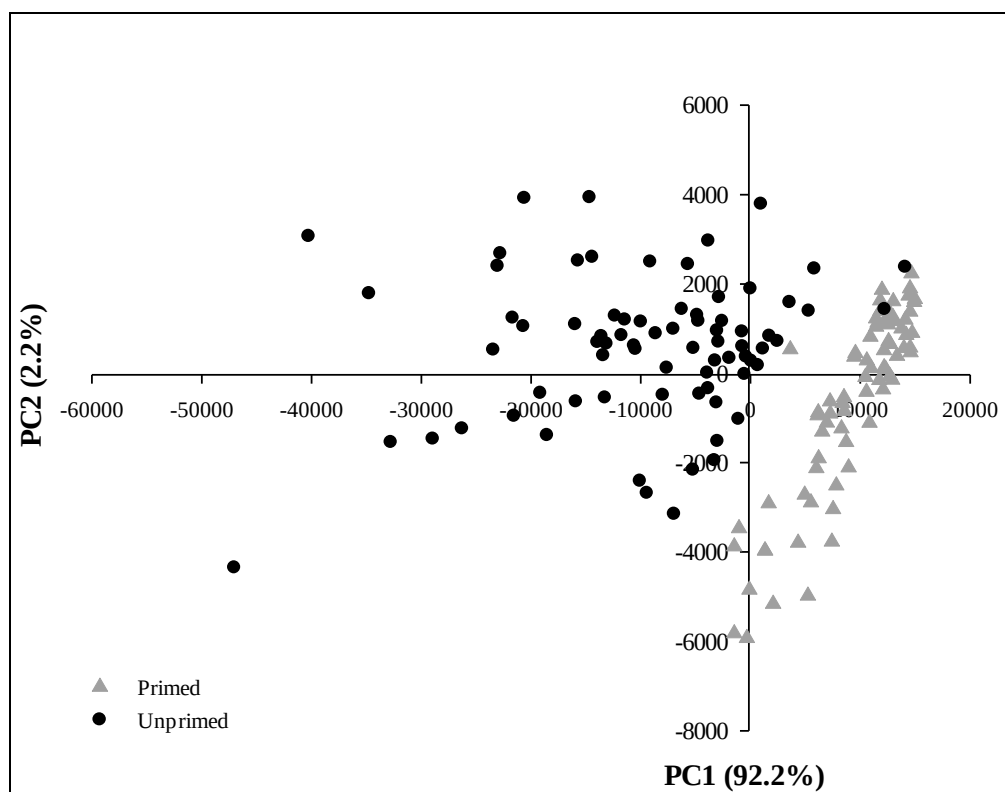


Fig. 4. Score plot of PCA performed with the LIBS spectra in primed and unprimed seeds (186 – 1042 nm).

Discrimination and classification methods

Discrimination and classification models

The discrimination and classification models were applied based on the scores of the two components, which explained 94.4% of the total variance. In general, the four models performed excellently in classifying primed and unprimed samples, as precision, accuracy, Kappa, F1, sensitivity and specificity were above 90% (Table 1). However, among these models, quadratic discriminant analysis (QDA) stood out among the others for its best accuracy (96%) and F1 score (97%). Accuracy, Kappa and specificity showed similar results with MDA with values of 97%, 94% and 96%, respectively. In the LDA model, 100% sensitivity was obtained, but in other parameters the percentages were lower compared to QDA. Therefore, QDA was considered the best model to classify the samples in the dataset.

Table 1. Performance parameters of the discrimination and classification methods of primed and unprimed seeds, based on dimensionality reduction scores by principal component analysis (PCA).

Model	Precision	Accuracy (%)	Kappa (%)	F1	Sensitivity (%)	Specificity (%)
LDA	91.93	95.28	90.6	95.61	100	90.74
QDA	96.30	97.17	94.34	97.2	98.14	96.22
MDA	96.29	97.17	94.34	97.17	98.08	96.22
KNN	93.08	95.28	90.55	95.44	98.08	92.38

Note: LDA, linear discriminant analysis; QDA, quadratic discriminant analysis; MDA, multiple discriminant analysis; KNN, k-nearest neighbor.

Figure 5. shows the confusion matrices obtained by the cross-validation tests performed for each of the classification models. The predictions were observed, in the first horizontal line true positive and false positive, in the second horizontal line the false negatives and false true, and on the diagonal the percentages of sensitivity and specificity (Vujović, 2021). These last two parameters are important in evaluating the classification of models.

The models that best classified the seed groups were QDA and KNN, where they correctly predicted primed seeds 22 times, that is, 22 seeds that were from this group were classified in the same one (Fig. 5B). The same happened with the unprimed ones, where they also were classified 22 times correctly (Fig. 5D). Additionally, primed and unprimed seeds were predicted 0 times incorrectly. In the other two LDA and MDA models, the classification fell in the unprimed group by two and one prediction, respectively (Fig. 5A and 5C).

The sensitivity and specificity for the QDA and KNN models were both 100% (Fig. 5B and 5D), however for the other LDA it was 91.7% and 95.65% for the MDA (Fig. 5A and 5C), having a decrease in the group of the prepared seeds.

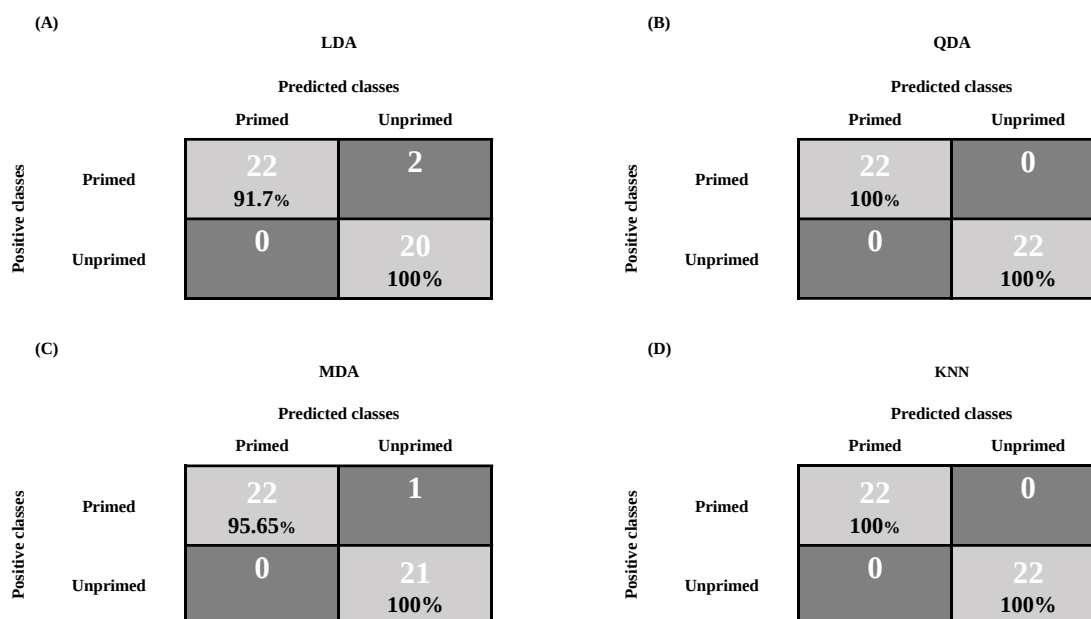


Fig. 5. Confusion matrix of discrimination and classification models. LDA, linear discriminant analysis (A); QDA, quadratic discriminant analysis (B); MDA, multiple discriminant analysis (C); KNN, k-nearest neighbors (D).

Biplot Analysis

The biplot of the areas of the elements with the highest intensity and the physiological variables of the seeds (PC1 = 52.2% and PC2 = 23.3%) explains 75.5% of the total data variability, therefore, PC1 and PC2 are sufficient to contribute with the variation of all data. Once again a two-group distinction was observed (Fig. 6). The primed samples were located in quadrants I and IV and have a relationship with the variables KI and KII and G5, which suggests that this group of seeds contains a higher potassium content, began their germination process on five days and took less time in reaching 50% of t50 germination (inverse relationship). The opposite happened with the unprimed seeds, shown in quadrants II and III, which contain more calcium (CaI and CaII),

magnesium (Mg), need more time to reach 50% of germination and did not show germination at day five, thus being a group of less vigorous seeds.

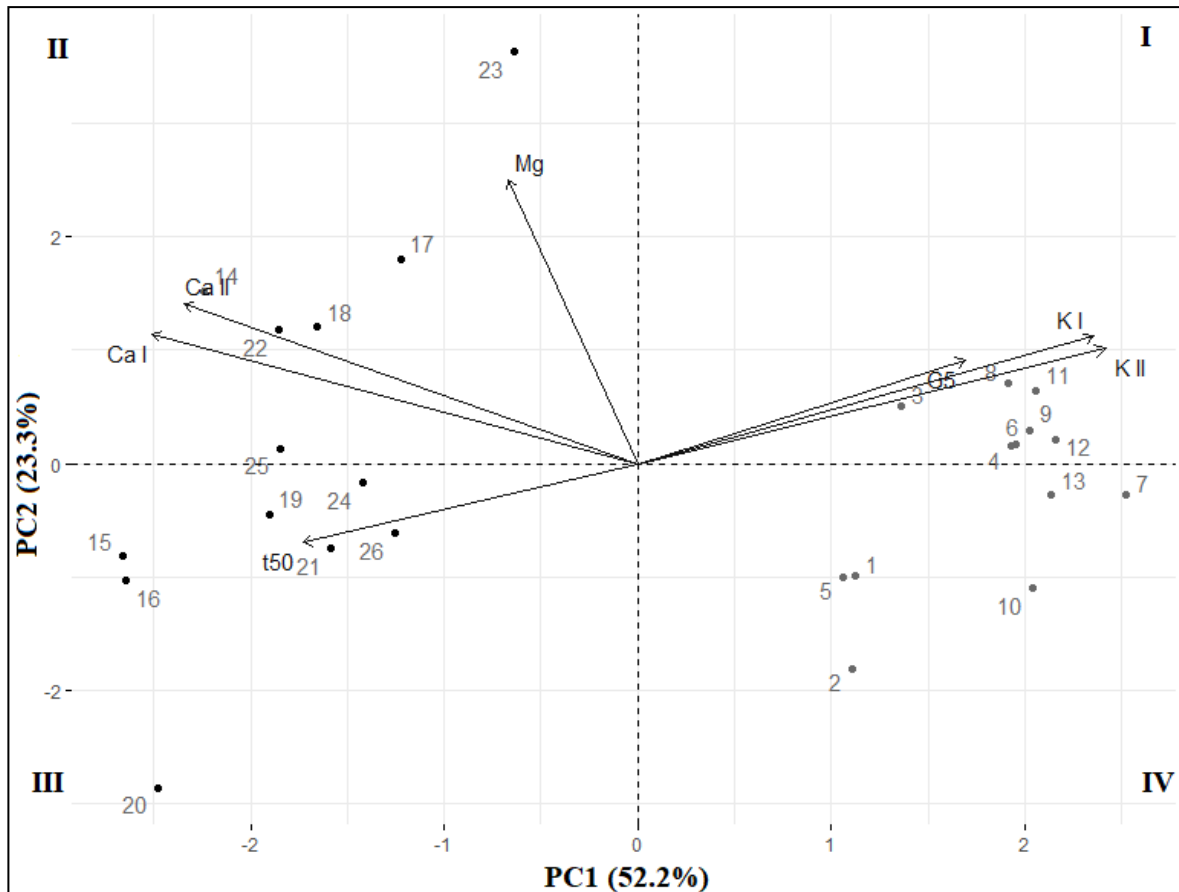


Fig. 6. Biplot component 1 and component 2 with scores and loadings.

2.4 Discussion

The priming protocol developed to be efficient should show positive effects on seed quality. The intended effect of priming is achieved by the greater uniformity and vigor of the primed seeds. However, these tests require work and time. In this study, we present an innovative solution capable of distinguishing primed and unprimed tomato (*S. lycopersicum*) seeds based on their elemental composition detected by Laser-induced breakdown spectroscopy (LIBS), which combined with principal component analysis and classification models such as linear, quadratic, multiple discriminants analysis (LDA, QDA,

MDA), and nearest neighbor (KNN) methods were able to accurately separate primed seeds from unprimed seeds.

Our results allowed us to observe the effect of priming with an increase in germination speed (Fig. 2A). Osmopriming produced seeds with superior physiological quality of *S. Lycopersicum* verified through the first count, the t50 percentage curve (Fig. 1B and 2A). It was confirmed that physiological conditioning allows fast germination to be achieved, which may be related to the activation of the metabolism of individual seeds with different potentials, obtaining uniformity. This process was also verified in previous studies carried out by Batista et al. (2020) and Petronilio et al. (2021). The LIBS technique was able to distinguish primed and unprimed tomato seeds based on their elemental composition. Mainly, it was observed that the spectral emission signal of the chemical element potassium in the primed seeds was higher, compared to the seeds that were not primed (Fig. 3).

The use of the LIBS technique and PCA revealed two groupings that showed the effectiveness of reducing the dimensionality of the data and the possibility of using the new data in classification models (Fig. 4).

In a study by Larios et al., 2020, the same technique was used, however there was no clear separation of the soybean seed lots with low and high vigor, and less explanation of the total variability of the data, differently from what was found in this study. This may be due to the priming and shape of the seeds, creating homogeneity in the tomato seeds, which allowed a better capture of the focusing lenses in the energetic pulse. Therefore, it is possible to infer that the average LIBS spectra combined with PCA in primed and unprimed tomato seeds can be used to reduce and distinguish the groups.

As seen in figure 4, unprimed seeds showed greater heterogeneity and more dispersed data, probably because of the difference in the spectral lines of the chemical elements. The opposite happened with the primed seeds, where the priming provided a certain uniformity. Therefore, it is important to highlight that this entire range of the average spectra of chemical elements, in particular potassium, calcium and magnesium, helped to highlight the differentiation of the groups, as they allowed the QDA model to classify primed and unprimed seeds with precision, sensitivity and specificity. (Table 1 and Fig. 5).

The elemental composition found in seeds is possibly related to phosphorus compounds through phytin, which is the main form of phosphate and mineral ions deposited in seeds during development, and mobilized during germination (Marcos-Filho, 2015). It is an insoluble salt of hexaphosphoric acid, made of Ca, K and Mg, which retains relatively large amounts of phosphorus. However, priming provided greater representation of the mineral K and consequently improved vigor (G5), compared to seeds that were not osmoprimed.

Calcium was representative in seeds that did not undergo priming and showed lower vigor. This result was consistent with that found by Larios et al. (2020), where in the discrimination of high and low vigor soybean seeds lots, it was discussed that calcium is an element that plays an important enzymatic role in the activation of the coenzyme Ca-calmodulin. Without the calcium, this enzyme does not have catalytic activity. The coenzyme Ca-calmodulin is necessary to activate several other enzymes, such as phospholipases, nucleotides, membrane Ca^{2+} ATPase, glutamate decarboxylase and NAD^+ kinase, and also to synthesize α -amylase that favors the starch breakdown in the germination of seeds.

In nine samples of edible seeds, in addition to the elements found in this study, phosphorus (P) and some minor elements such as sulfur (S) and zinc (Zn) were also identified. The elements P and Mg showed a positive correlation, however the correlation can be influenced depending on the quality of the seeds (Gamela et al., 2019). In dry bean grains of different species, it was found high concentrations of the elements K, Mg and P, and LIBS was considered a fast analytical way that contributes to the properties of seeds and sustainable environments, since this method is not destructive (Gamela et al., 2020). Chromium (Cr), a different element to those in this study, was detected in corn grains (Jabbar et al., 2020).

Cross-validation indicates that primed tomato seeds can be successfully differentiated from unprimed ones with 100% sensitivity and specificity using the QDA model. The simple preparation, the sample quantity and the capacity for direct and rapid analysis demonstrated the potential of LIBS as a new methodology to be used in detecting priming in tomato seeds as an alternative to traditional methodologies used by the seed industry. Therefore, with this work we propose for the first time the use of the induced

burst spectroscopy (LIBS) technique combined with PCA and classification models to accurately select and distinguish primed and unprimed tomato seed samples. The chemical elements that serve as the main classification between the two groups of samples are potassium (K), calcium (Ca) and magnesium (Mg).

References

Abdi H, Williams LJ, and Valentin D (2013) Multiple fator analysis: Principal componente analysis for multitable and multiblock data sets. *WIREs Computational Statistics* **5**, 149 – 179.

Batista TB, Binotti FFS, Cardoso ED, Bardivieso EM, Costa E (2015) Aspectos fisiológicos e qualidade de mudas da pimenteira em resposta ao vigor e condicionamento das sementes [Physiological aspect sand quality of pepper seedlings in response to vigor and conditioning seeds]. *Bragantia* **74**, 367–373.

Batista TB, Fernandez GJ, da Silva TA, Maia J and Silva EAA (2020) Transcriptome analysis in osmo-primed tomato seeds with enhanced longevity by heat shock treatment. *AoB Plants* **12**, 1-10.

BRASIL (2009) Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília, MAPA/ACS, p 395.

Costa VC, Augusto AS, Castro JP, Machado RC, Andrade DF, Babos DV, Sperança MA, Gamela RR and Pereira-Filho ER (2019). Espectroscopia de quebra induzida por laser (LIBS): histórico, fundamentos, aplicações e potencialidades. *Química Nova* **42**, 527-545.

Castro J P and Pereira-Filho ER (2016) Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS). *Journal of Analytical Atomic Spectrometry* **31**, 2005-2014.

Ferreira DDS, Rodrigues LDS, Pereira FMV and Pereira e Pereira-Filho ER (2023) Principal component analysis (PCA) para a avaliação de dados químicos e geração de heat maps: Um tutorial. *Quimica Nova* **XY**, 1-8.

Gamela RR, Sperança MA, Andrade DF, and Pereira-Filho ER (2019) Hyperspectral images: a qualitative approach to evaluate the chemical profile distribution of Ca, K, Mg, Na and P in edible seeds employing laser-induced breakdown spectroscopy. *Analytical Methods* **11**, 5543-5552.

Gamela RR, Costa VC, Sperança MA and Pereira-Filho ER (2020) Laser-induced breakdown spectroscopy (LIBS) and wavelength dispersive X-ray fluorescence (WDXRF) data fusion to predict the concentration of K, Mg and P in bean seed samples. *Food Research International* **132**, 109037.

Habibi N, Aryan S, Amin MW, Sanada A, Terada N, Koshio K (2023) Potential Benefits of Seed Priming under Salt Stress Conditions on Physiological, and Biochemical Attributes of Micro-Tom Tomato Plants. *Plants* **12**, 2187.

Hossin M and Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. *Internacional Journal of Data Mining & Knowledge Management Process* **5**, 1-11.

Jabbar A, Akhtar M, Mahmood S, Ahmed R and Baig MA (2020) Determination of major inorganic nutrients in maize tissues by calibration-free laser induced breakdown spectroscopy. *Analytical Letters* **53**, 1328-1341.

Machado RC, Andrade DF, Babos DV, Castro JP, Costa VC, Sperança MA, Garcia JA, Gamela RR, Pereira-Filho ER (2020) Solid sampling: advantages and challenges for chemical element determination—a critical review. *Journal of Analytical Atomic Spectrometry* **35**, 54-77.

Mauro RP, Rizzo V, Leonardi C, Mazzaglia A, Muratore G, Distefano M, Sabatino L, Giuffrida F (2020) "Influence of harvest stage and rootstock genotype on compositional and sensory profile of the elongated tomato cv. "Sir Elyan". *Agriculture* **10**, 82.

Verma AK, Tiwari R, Karthik K, Chakraborty S, Deb R and Dhama K (2013) Nutraceuticals from fruits and vegetables at a glance: A review. *Journal of Biological Sciences* **13**, 38.

Mingoti SA (2005) Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte, UFMG.

Marcos-Filho J (2015) Fisiologia de sementes de plantas cultivadas.

Noll R and Noll R (2012) Laser-Induced Breakdown Spectroscopy. Fundamentals and Applications, Springer Berlim Heidelberg.

Larios GS, Nicolodelli G, Senesi GS, Ribeiro MC, Xavier AA, Milori DM, Alves CZ, Marangoni BS, Cena C (2020) Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishinhg High- and Low-Vigor Soybean Seed Lots. *Food Analytical Methods* **13**, 1691-1698.

Villela FA, Filho DL and Sequeira EL (1991) Tabela de potencial osmótico em função da concentração de polietileno glicol 6.000 e da temperatura. *Pesquisa Agropecuária Brasileira*, **26**, 1957-1968.

Vujović Ž (2021) Classification model evaluation metrics. *Internacional Journal of Advanced Computer Science and Applications* **12**, 599-606.

Pasquini C, Cortez J, Silva LMC, Gonzaga FB (2007) Laser Induced Breakdown Spectroscopy. *J. Braz. Chem. Soc* **18**, 463-512.

Petronilio ACP, Batista TB, da Silva EAA. Osmo-priming in tomato seeds down-regulates genes associated with stress response and leads to reduction in longevity. *Seed Science Research* **31**, 211-216.

Pořízka P, Klus J, Képeš E, Prochazka D, Hahn DW, Kaiser J (2018) On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review. *Spectrochim Acta - Part B* **148**, 65–82.

CONSIDERAÇÕES FINAIS

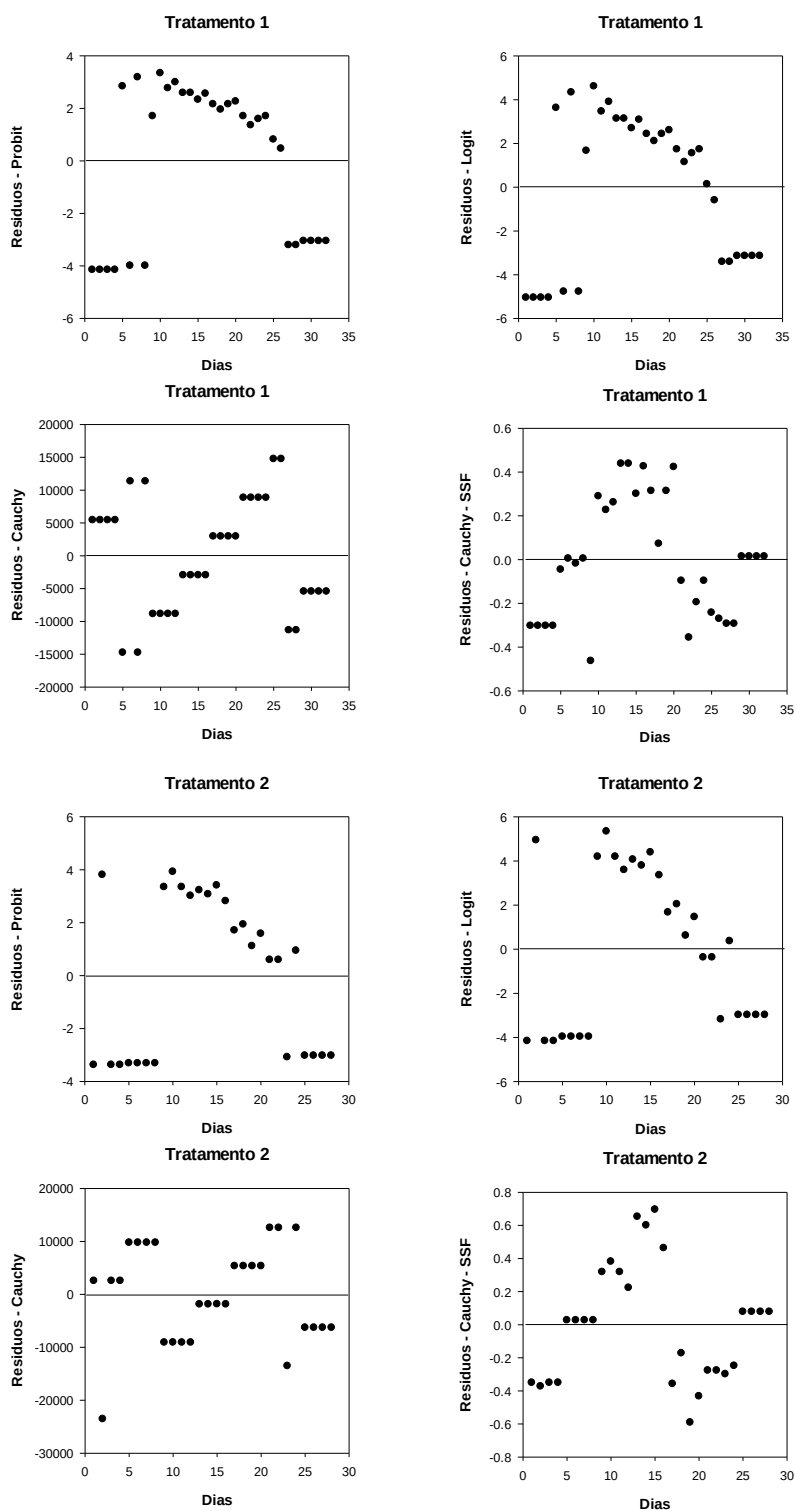
Atualmente, a agricultura e o setor sementeiro enfrentam grandes desafios na produção e conservação das espécies, especialmente de produtos que contêm fontes de vitaminas essenciais para a segurança alimentar, como o tomate. Aqui, foi apresentada uma maneira robusta para prever a longevidade de sementes de tomate utilizando a função de ligação Cauchy-SSF com o método de mínimos quadrados ordinários (OLS), que superou o Probit. Esse conhecimento pode ser de grande utilidade para os bancos de germoplasma, auxiliando na renovação de materiais e na tomada de decisões assertivas frente à conservação da espécie.

Foi igualmente relevante descobrir que a técnica de espectrometria de emissão óptica com plasma induzido por laser (LIBS), combinada com a estatística multivariada e modelos de classificação, seleciona de maneira precisa sementes de tomate primed e não primed, contribuindo para o avanço da ciência e da tecnologia de sementes na indústria brasileira. Indiretamente, esses benefícios se estendem à Colômbia, resultando em uma melhoria na qualidade das sementes de tomate. É sabido que o Brasil é o terceiro maior fornecedor de sementes de tomate para a Colômbia, logo após os Estados Unidos e China.

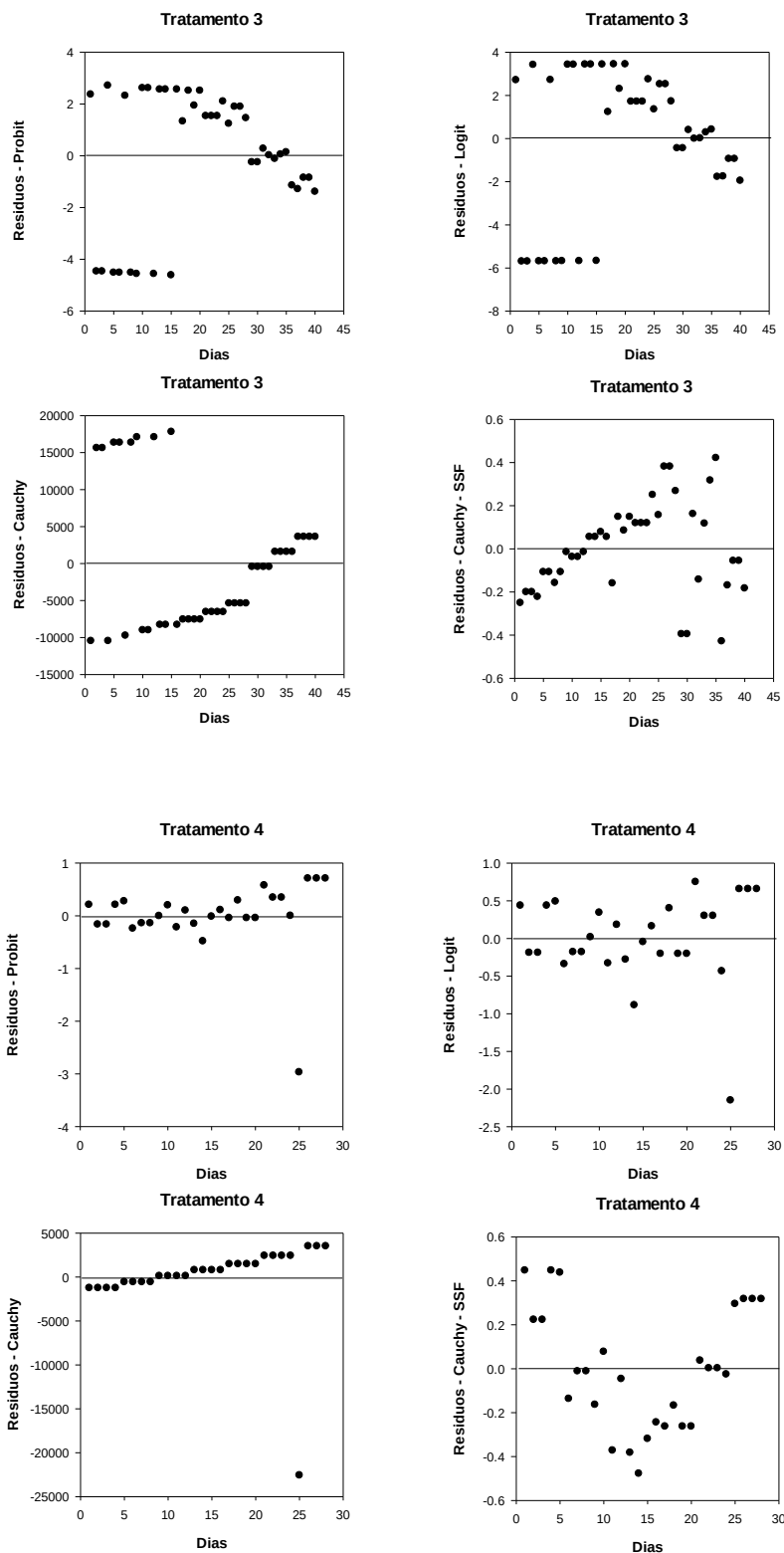
REFERÊNCIAS

- BATISTA, T.B. **Longevidade de sementes condicionadas de tomate: estudos fisiológicos e moleculares**. Dissertação (Mestre em Agronomia - Agricultura) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2018.
- BERNAL-LUGO, I and; CARL LEOPOLD, A. Changes in soluble carbohydrates during seed storage. **Plant Physiology**, v. 98, n. 3, p. 1207–1210, 1992.
- CONAB. **Companhia Nacional de Abastecimento**. Tomate: Análise dos indicadores da produção e comercialização no mercado mundial, Brasileiro e Catarinense, v. 21, 2019. Disponível em: <https://www.conab.gov.br/institucional/publicacoes/compendio-de-estudos-da-conab/item/12529-compendio-de-estudos-da-conab-v-21-tomate-analise-dos-indicadores-da-producao-e-comercializacao-no-mercado-mundial-brasileiro-e-catarinense>
- DE SOUZA, D. et al. Galactinol as marker for seed longevity. **Plant Science**, Ireland, v. 246, p. 112–118, 2016.
- HABIBI, N., ARYAN, S., AMIN, M. W., SANADA, A., TERADA, N., & KOSHIO, K. Potential Benefits of Seed Priming under Salt Stress Conditions on Physiological, and Biochemical Attributes of Micro-Tom Tomato Plants. **Plants**, v. 12, n. 11, p. 2187, 2023.
- HAY, F. R. et al. Seed longevity phenotyping: recommendations on research methodology. **Journal of experimental botany**. v. 70, n. 2, p. 425–434, 2019.
- MINAGRICULTURA. **Ministerio de Agricultura y Desarrollo Rural de Colombia**. Área, Producción y Rendimiento Nacional por Cultivo en Colombia, 2018. Disponível em: <https://www.agronet.gov.co/estadistica/Paginas/home.aspx?cod=1>.
- RAJJOU, L.; DEBEAUJON, I. Seed longevity: Survival and maintenance of high germination ability of dry seeds. **Comptes Rendus - Biologies**, v. 331, n. 10, p. 796–805, 2008.
- THWE, A. A. et al. Impact of red and blue nets on physiological and morphological traits, fruit yield and quality of tomato (*Solanum lycopersicum* Mill.). **Scientia Horticulturae**, Netherlands, v. 264, p. 109185, 2020.

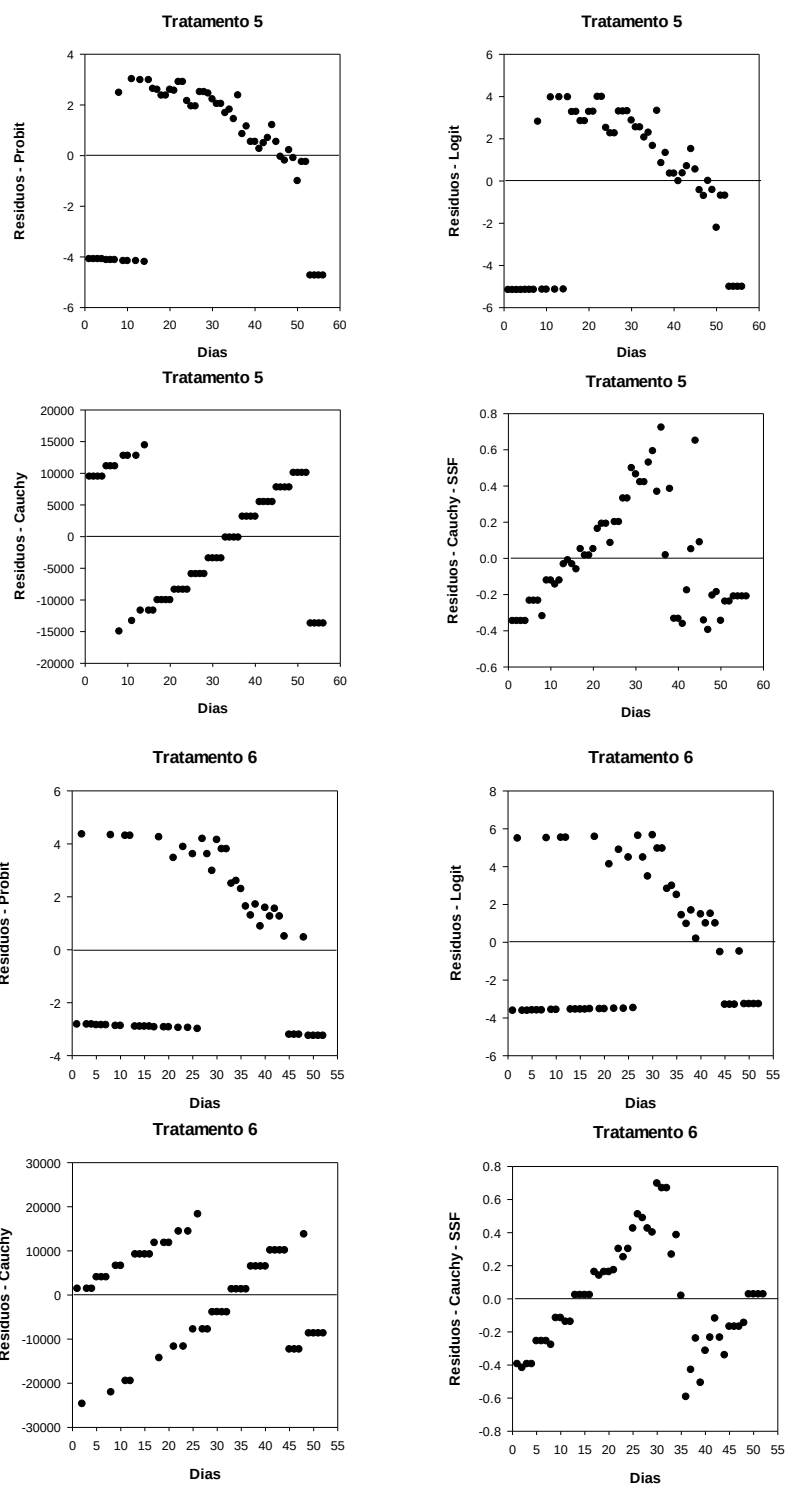
APÊNDICE A – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 1 e 2.



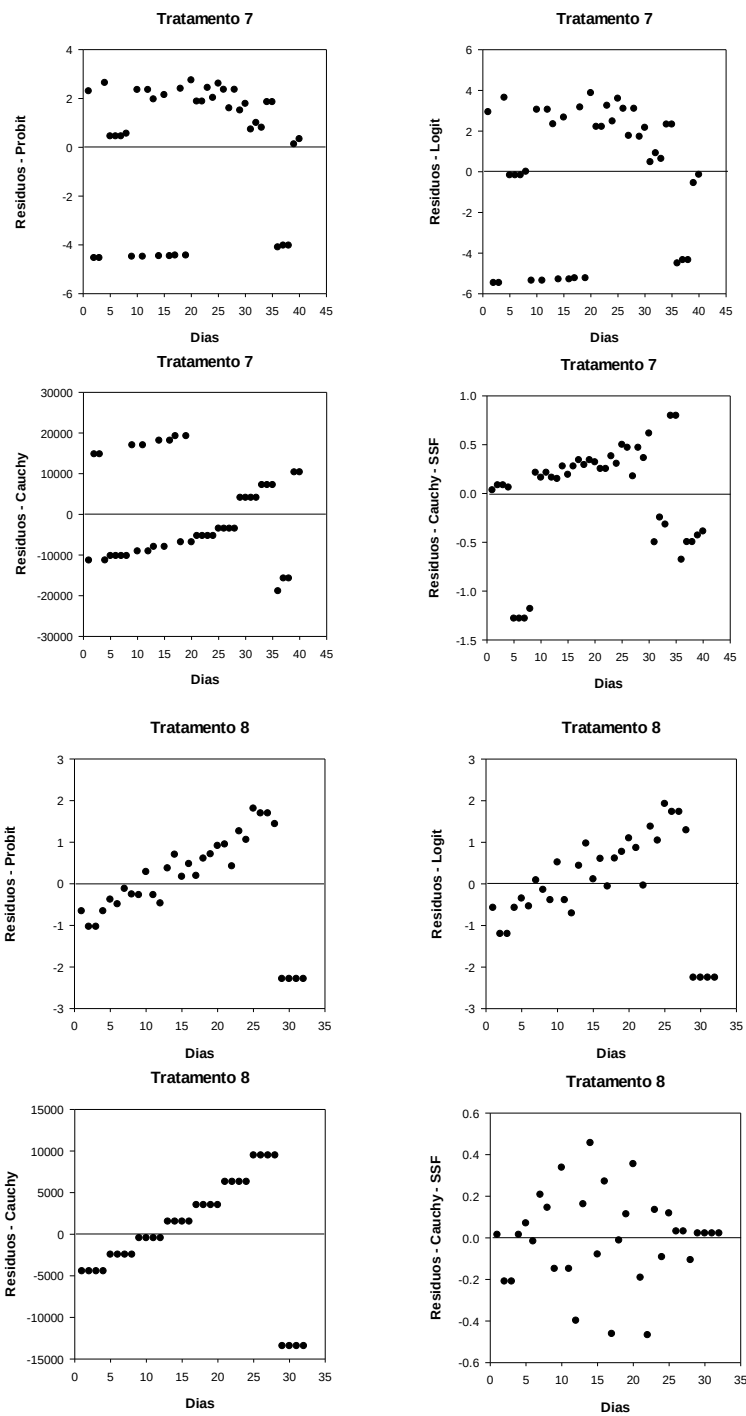
APÊNDICE B – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 3 e 4.



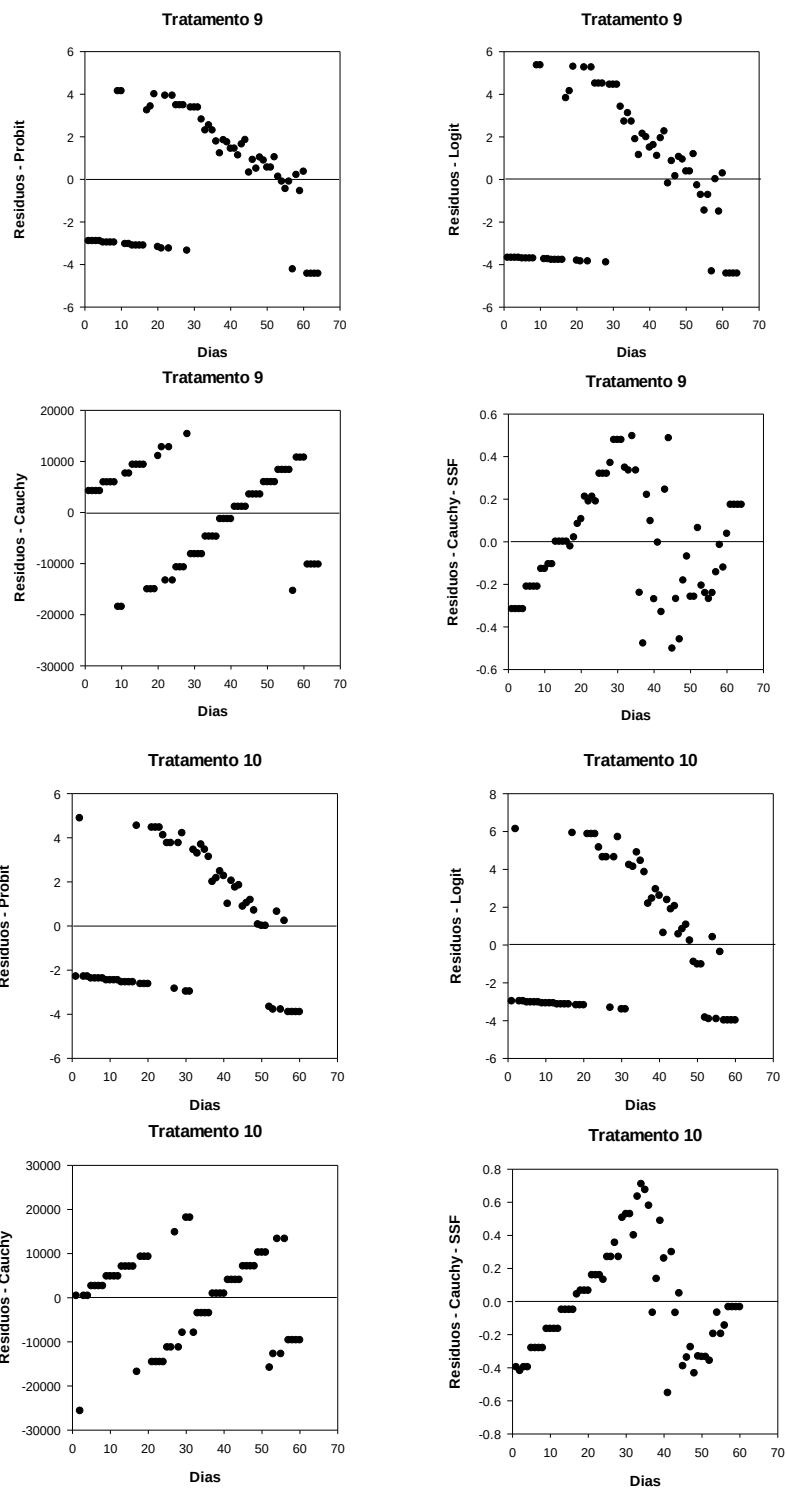
APÊNDICE C – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 5 e 6.



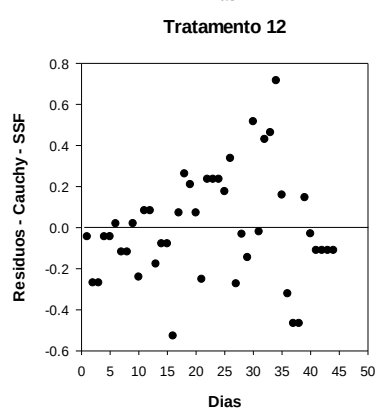
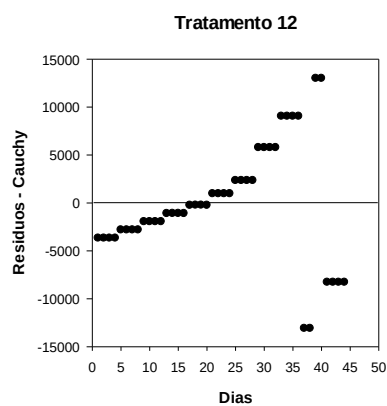
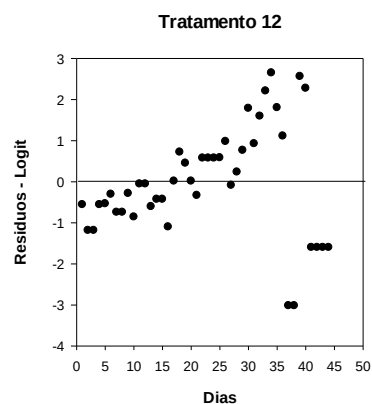
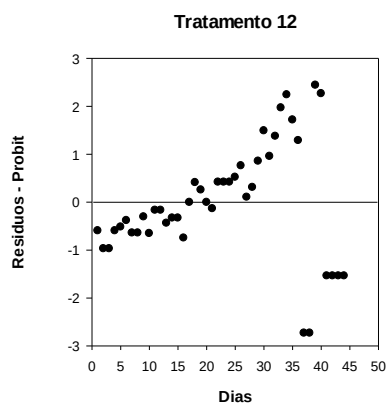
APÊNDICE D – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 7 e 8.



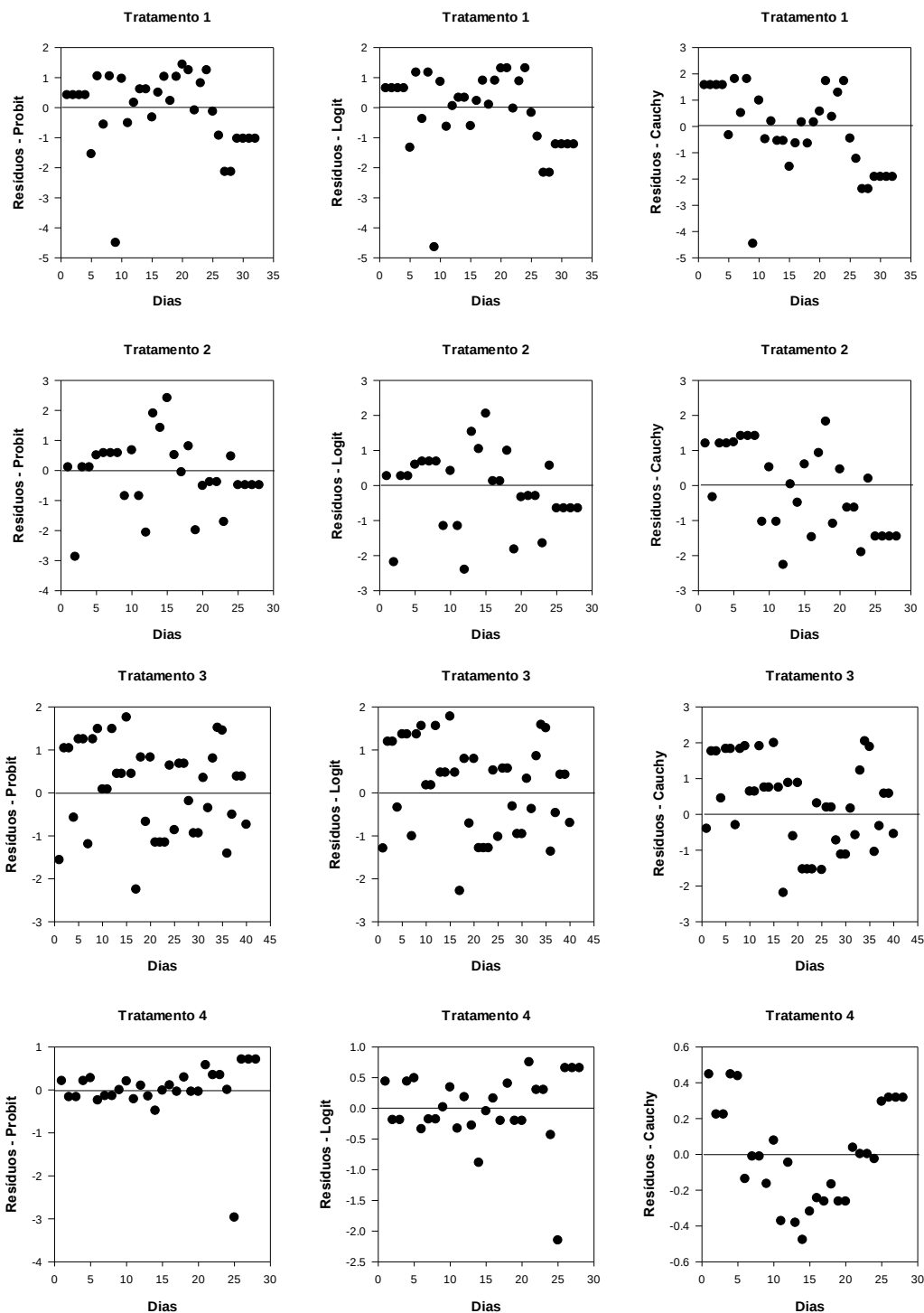
APÊNDICE E – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 9 e 10.



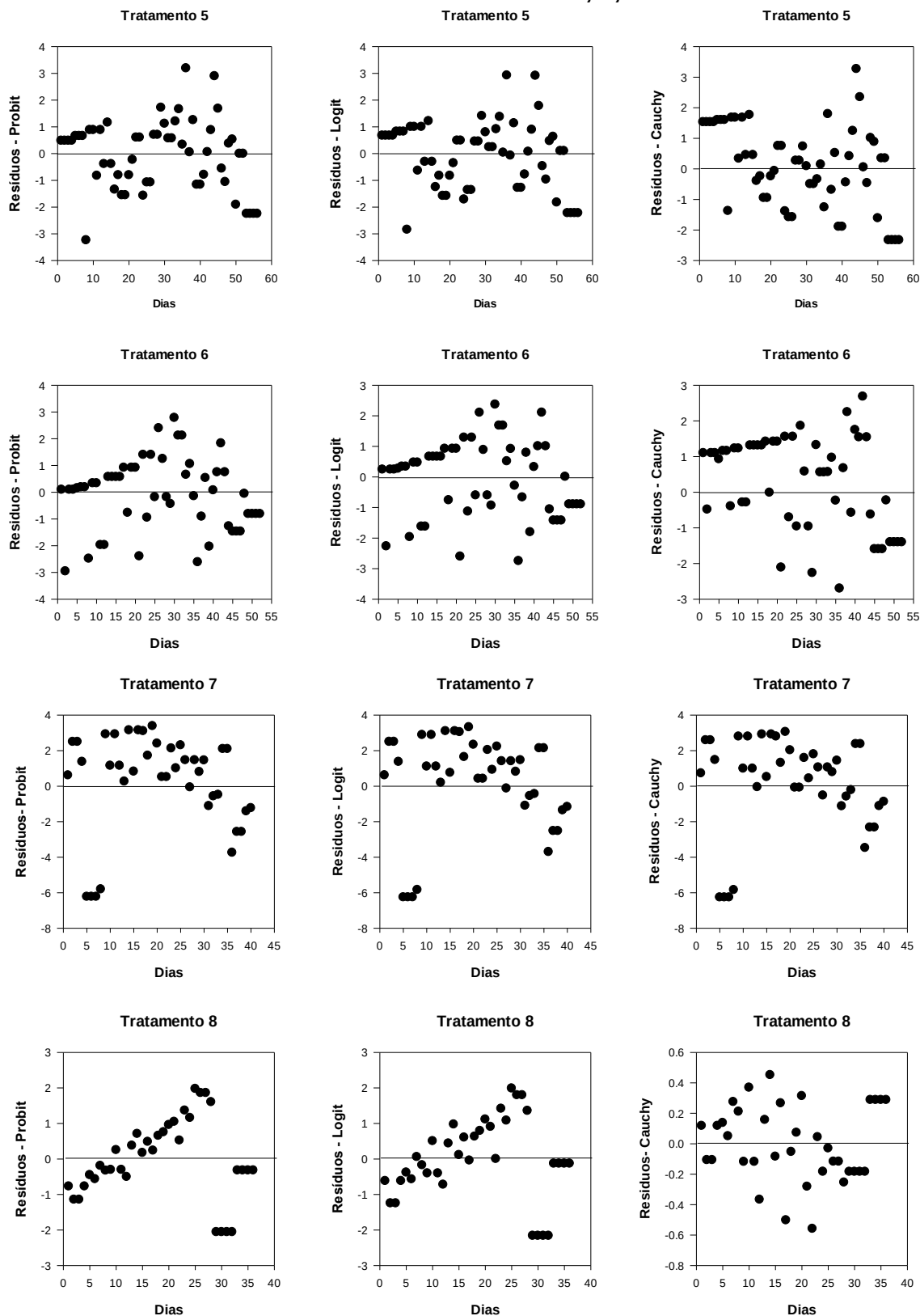
APÊNDICE F – Resíduos das funções Probit, Logit, Cauchy e Cauchy - SSF pelo método OLS nos tratamentos 9.



APÊNDICE G – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 1, 2, 3 e 4.



APÊNDICE H – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 5, 6, 7 e 8.



APÊNDICE I – Resíduos das funções Probit, Logit e Cauchy pelo modelo GLM nos tratamentos 9, 10, 12.

