LUCAS DE PAULO ARCANJO

SAMPLING PLANS AND APPLICATION OF NEURAL NETWORKS TO FORECAST THE SEASONAL DYNAMICS OF *Bemisia tabaci* IN SOYBEAN CROPS

Thesis submitted to the Entomology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Marcelo Coutinho Picanço

VIÇOSA - MINAS GERAIS
2023

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

T

A668s
2023

Arcanjo, Lucas de Paulo, 1994-
Sampling plans and application of neural networks to forecast the seasonal dynamics of *Bemisia tabaci* in soybean crops: / Lucas de Paulo Arcanjo. – Viçosa, MG, 2023.
1 tese eletrônica (76 f.): il. (algumas color.).

Texto em inglês.
Orientador: Marcelo Coutinho Picanço.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Entomologia, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.453
Modo de acesso: World Wide Web.

1. *Bemisia tabaci. 2.* Mosca-branca - Controle. 3. Soja - Doenças e pragas. 4. Amostragem. 5. Redes neurais (Computação). I. Picanço, Marcelo Coutinho, 1958-. II. Universidade Federal de Viçosa. Departamento de Entomologia. Programa de Pós-Graduação em Entomologia. III. Título.

CDD 22. ed. 595.754

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

LUCAS DE PAULO ARCANJO

SAMPLING PLANS AND APPLICATION OF NEURAL NETWORKS TO FORECAST THE SEASONAL DYNAMICS OF *Bemisia tabaci* IN SOYBEAN CROPS

Thesis submitted to the Entomology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: July 8, 2023.

Assent:

gov.br
Documento assinado digitalmente
LUCAS DE PAULO ARCANJO
Data: 03/08/2023 22:01:06-0300
Verifique em https://validar.iti.gov.br

Lucas de Paulo Arcanjo
Author

gov.br
Documento assinado digitalmente
MARCELO COUTINHO PICANCO
Data: 04/08/2023 12:43:30-0300
Verifique em https://validar.iti.gov.br

Marcelo Coutinho Picanço
Adviser

A Deus, à minha esposa, aos meus pais e irmão.

AGRADECIMENTOS

A Deus.

À minha esposa (Bianca) por toda parceria e carinho.

Aos meus pais (Vicente e Célia) e irmão (Gabriel) por toda orientação, instrução e apoio

A todos os membros do laboratório de Manejo Integrado de Pragas (MIP) da Universidade Federal de Viçosa, no qual tive todo o apoio, aprendizado e amizade desde o início da graduação.

Ao meu orientador Marcelo Coutinho Picanço e a sua família por toda a orientação, direcionamento e amizade desde o início da graduação.

À Universidade Federal de Viçosa, pela oportunidade de concluir o curso de pós-graduação.

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código Financeiro 001.

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

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa.

“Você ganha força, coragem e confiança através de cada experiência em que você realmente para e encara o medo de frente.” (Eleanor Roosevelt)

ABSTRACT

ARCANJO, Lucas de Paulo, D.Sc., Universidade Federal de Viçosa, July, 2023. **Sampling plans and application of neural networks to forecast the seasonal dynamics of *Bemisia tabaci* in soybean crops.** Adviser: Marcelo Coutinho Picanço.

Soybean *Glycine max* (L) (Merr) is the most produced, consumed, and traded legume worldwide. After its establishment in Brazil, *Bemisia tabaci* became a notorious sucking pest on soybean. Robust sampling plans and seasonal dynamic studies of *B. tabaci* in tropical soybean areas are essential to technicians and farmers early detect pest populations and plan sprays to manage this pest on time. The aim of this study is to determine a sampling plan and seasonal dynamics of *B. tabaci* in soybean crops through artificial neural networks. These studies were carried out in soybean commercial fields. Whitefly density, climatic elements and soybean age were assessed to support the dataset. In the seasonal dynamic studies, artificial neural networks were developed and selected to study this pest dynamic. The sampling design in this study is composed of 49 samples. The sampling unit and technique are the apical part of the soybean canopy and beating a plastic tray against plant apex, respectively, throughout the plant stages. The artificial neural network structure selected to determine the seasonal dynamic of *B. tabaci* in soybean crops has five entries (soybean age, average temperature, rainfall, wind speed, and atmosphere pressure) and four neurons in the hidden shell. This model previews whitefly adults with high accuracy from seven days of lag; it is reliable for modelling the seasonal dynamics of the whitefly *B. tabaci* in soybean crops. In conclusion, this study provides technical tools to scout, early detect, and plan sprays against the whitefly population, avoiding pest outbreaks.

Keywords: Sucking pests. AI. Conventional sampling plan. Forecast.

RESUMO

ARCANJO, Lucas de Paulo, D.Sc., Universidade Federal de Viçosa, julho de 2023. **Planos de amostragem e aplicação de redes neurais para previsão da dinâmica sazonal de *Bemisia tabaci* em lavouras de soja.** Orientador: Marcelo Coutinho Picanço.

A soja *Glycine max* (L) (Merr) é a leguminosa mais produzida, consumida e comercializada no mundo. Após seu estabelecimento no Brasil, *Bemisia tabaci* tornou-se uma praga sugadora notória da soja. Planos de amostragem robustos e estudos de dinâmicas sazonais de *B. tabaci* em áreas tropicais de soja são essenciais para que técnicos e agricultores detectem precocemente populações de pragas e planejem pulverizações para manejar essa praga a tempo. O objetivo deste estudo é determinar um plano de amostragem e dinâmica sazonal de *B. tabaci* em lavouras de soja por meio de redes neurais artificiais. Esses estudos foram realizados em campos comerciais de soja. Densidade de mosca branca, elementos climáticos e idade da planta de soja foram avaliados para compor o conjunto de dados. Nos estudos de dinâmica sazonal, uma rede neural foi desenvolvida e selecionada para estudar a dinâmica dessa praga. O plano de amostragem deste estudo é composto por 49 amostras. A unidade amostral e a técnica são a parte apical do dossel da soja e batida de bandeja plástica contra o ápice da planta, respectivamente, ao longo das fases da planta. A estrutura da rede neural selecionada para determinar a dinâmica sazonal de *B. tabaci* na cultura da soja possui cinco entradas (idade da planta de soja, temperatura média, precipitação, velocidade do vento e pressão atmosférica) e quatro neurônios na camada oculta. Este modelo prevê os adultos da mosca branca com alta precisão com sete dias de antecedência e é confiável para modelar a dinâmica sazonal da mosca branca *B. tabaci* em lavouras de soja. Em conclusão, este estudo fornece ferramentas técnicas para monitorar, detectar precocemente e planejar pulverizações contra mosca-branca, evitando surtos de pragas.

Palavras-chave: Pragas sugadoras. IA. Plano de amostragem convencional. Previsão.

SUMMARY

1. General Introduction

Soybean *Glycine max* (L) (Merr) is the leading oilseed traded worldwide (Foreign Agricultural Service, 2023). In the season 2022-2023, 96.59 million tons were exported (Foreign Agricultural Service, 2023). The soybean world production was 388 million tons in 2021 on 137.9 million ha across the world (FAOSTAT, 2021). The top seven soybean producers in the world are respectively Brazil, the United States of America, Argentina, China, India, Paraguay, and Canada, being responsible for 93% of the world's production (Foreign Agricultural Service, 2023). Nowadays, *G. max* crops are receiving technological imputes from planting to harvest, and production of 6000 kg per ha has been achieved in well-managed areas in Brazil (Agrolink, 2022). However, soybean cost production is increasing while this commodity price is unstable, soybean price was $566 per ton in February 2023 and decrease to $530 per ton in March (6.4% off) (Foreign Agricultural Service, 2023). Thus, protecting soybean against pathogens, insects, and weeds are essential to reach the maximum plant yield.

The whitefly *Bemisia tabaci* (Gennadius 1889) (Hemiptera: Aleyrodidae) remains an issue in soybean crops, especially in dry and hot areas across tropical lands (Padilha et al., 2021; Schutze, Naranjo, and Yamamoto, 2022). This generalist pest sucks phloem sap and injects toxins into over 1000 plants worldwide (Jones, 2003; Li et al., 2021). Also, whitefly transmits viruses causing economic diseases to essential crops (Jones, 2003; Li et al., 2021). In a short period of time, the whitefly population increase, it can run three cycles in a soybean crop and promote damage above 1000 kg ha^{-1}, besides reducing protein content in 440 kg. ha^{-1} (Schutze, Naranjo, and Yamamoto, 2022). In areas with high pest pressure historic, scouting and controlling this pest before soybean rows close (35-40 days after emergence) is mandatory to avoid losses. Therefore, early detection and monitoring are essential to manage whitefly in soybean fields.

Sampling plans are technical methods developed to scout pests (Pedigo and Buntin, 1993). These plans evaluate pest populations with accuracy, precision, and must be feasible (Costa et al., 2019; Lopes et al., 2019; Mota et al, 2022 Neves et al., 2023). Sampling plans could be standardized or sequential. Standardized sampling plans are the first step to scout pests in Integrated Pest Management (IPM) programs. In this method, a pest population is estimated through fixed numbers of samples (Mota et al, 2022). Instead, sequential sampling plans are the second level of the sampling method where the numbers of samples necessary to take control or non-control vary according to a pest population level and decisions use to be faster (Costa et al., 2019; Lopes et al., 2019; Neves et al., 2023). There are no standardized sampling plans established to scout whitefly adults in tropical soybean crops.

Artificial intelligence (AI) is any technique that enables a computer to mimic humans' behavior (Amini, 2019). AI is mainly applied to solve regression, classification, dimensionality reduction, and clustering problems from data extracted of images, videos, audio, counting and so forth (Amini, 2019; Benos et al., 2021). The general idea of AI algorithms is training and validate a dataset to establish a model or rules able to predict or classify other data (supervised learning) or extract data without a dataset with independent (y) and dependent attributes (x) connected (unsupervised learning) (Benos et al., 2021; Prabhu, Parab, and Naik, 2021; Mendoza et al., 2023). Machine learning is a branch of AI that contains classic algorithms such as decision trees, support vector machine, and random forest e.g otherwise deep learning approaches extract information using neural networks (ANN) (Khan, R., Dhingra, N., & Bhati, N. 2022; Liakos et al., 2018; Oliveira et al., 2020).

Sampling plans and ANN would be interesting methods to estimate the whitefly population in soybean crops and modeling its seasonal dynamic. The seasonal dynamic of whitefly studies makes it possible to understand pest abundance over the years from models built upon biotic and abiotic features. Modeling the season dynamic of whitefly makes possibly understand principal features that govern its population dynamic and forecast it population. Then, this approach could be applied to preview and early detect pest populations helping farmers control whiteflies in time avoiding losses. Therefore, this thesis is composed by two chapters in scientific paper format. In chapter one, a standardized sampling plan to scout *B. tabaci* adults is shown; otherwise, chapter two studies the seasonal dynamic of whitefly in soybean fields through an artificial intelligence technique.

1.1 References

Agrolink, 2022. Accessed on May 2023. https://www.agrolink.com.br/colunistas/coluna/soja-com-alta-produtividade--mais-rentavel--mais-sustentavel_467380.html

Amini, A. (2019). Introduction to deep learning. MIT, 6, S191.

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Costa, T.L., Sarmento, R.A., de Araujo, T.A., Pereira, P.S., Silva, R.S., Lopes, M.C., & Picanco, M. C. (2019). Economic injury levels and sequential sampling plans for *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotype B on open-field melon crops. *Crop Protection*, 125, 104887.

Food and Agriculture Organization of United Nations. FAOSTAT, 2021.

Foreign Agricultural Service. Oilseeds: World Markets and Trade. USDA, 2023.

Jones, D.R. (2003). Plant viruses transmitted by whiteflies. *European journal of plant pathology*, *109*, 195–219.

Khan, R., Dhingra, N., & Bhati, N. (2022). Role of artificial intelligence in agriculture: A comparative study. In Transforming Management with AI, Big-Data, and IoT (pp. 73–83). Cham: Springer International Publishing.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*,*18*(8), 2674.

Lopes, M.C., Farias, E.S., Costa, T.L., Arcanjo, L.P., Santos, A.A., Ribeiro, A.V., ... & Picanco, M. C. (2019). Economic injury level and sequential sampling plan for *Liriomyza huidobrensis* management in tomato crops. *Crop Protection*, *124*, 104848.

Mendoza, Q.A., Pordesimo, L., Neilsen, M., Armstrong, P., Campbell, J., & Mendoza, P. T. (2023). Application of Machine Learning for Insect Monitoring in Grain Facilities. *AI*, *4*(1), 348–360.

Neves, D.V.C., Lopes, M.C., Pereira, P.S., Reis, K.H.D.B., Picanço, M.M., Costa, T. L., ... & Picanço, M. C. (2023). Sequential sampling plans for thrips (Thysanoptera: Thripidae) in soybean crops with insecticide applications using tractor or airplane. *Journal of Economic Entomology*, *116*(2), 599–604.

Oliveira Aparecido, L.E., Souza Rolim, G., Moraes, J.R.S.C., Costa, C.T.S., & Souza, P. S. (2020). Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, *64*, 671–688.

Pedigo, L.P., & Buntin, G.D. (Eds.). (1993). Handbook of sampling methods for arthropods in agriculture. CRC press.

Prabhu, Y. V., Parab, J.S., & Naik, G.M. (2021). Back-Propagation Neural Network (BP-NN) model for the detection of borer pest attack. In *Journal of Physics: Conference Series* (Vol. 1921, No. 1, p. 012079). IOP Publishing.

Schutze, I. X., Naranjo, S. E., & Yamamoto, P. T. (2022). Impact of *Bemisia tabaci* MEAM1 (Hemiptera: Aleyrodidae) on soybean yield and quality under field conditions. *Journal of Economic Entomology*, *115*(3), 757–766.

Souza, M.D.S.C., Lopes, M.C., Arcanjo, L.P., Mesquita, A. L. M., Pereira, R.D.C.A., Silva, R.S., & Picanço, M.C. (2022). Designing a robust strategy to scout a key pest in cashew orchards. *Crop Protection*, *157*, 105954.

2 Chapter 1. A novel standardized sampling plan to Bemisia tabaci (Hemiptera: Aleyrodidae) in neotropical soybean crops

2.1 Abstract

Soybean *Glycine max* (L) (Merr) is the most produced and consumed oilseed globally. After establishing in Brazilian fields, *Bemisia tabaci* (Gennadius 1889) (Hemiptera: Aleyrodidae) became a primary pest on soybean plantations. Methods to scout pests in Brazilian soybean fields come from investigations developed in North America, and they are not realistic for some pests in tropical agriculture. Regarding whitefly, a lack of knowledge on sampling methods in Neotropical soybean areas persists. A standardized sampling plan is the first step to track pest population in pest management programs. Hence, this study aimed to determine a novel standardized sampling protocol to scout *B. tabaci* adults in tropical agriculture. This study was performed in 26 commercial soybean fields at vegetative, flowering, and fruiting stages in Brazilian Savanah from 2017 to 2021. Whitefly adults were evaluated to establish a standardized sampling plan following five steps: (i) sample unit determination, (ii) sample technique evaluation, (iii) frequency distribution of whitefly density, (iv) the optimum number of samples calculation, and (v) costs and time of sampling process. Thus, the ideal sampling unit to evaluate whitefly through the plant cycle was the apical part of the soybean canopy. The most appropriate sampling technique was beating the apical site of the plant against a plastic tray. The number of plants demanded to scout this pest was 49 samples per field. The timing (maximum 5 minutes per ha) and costs (maximum of US$ 0.74 per ha) of this sampling method met the soybean operations requirement. Therefore, this protocol is precise and representative and has a reasonable cost suitable for monitoring *B. tabaci* in tropical soybean fields.

Keywords: Aleyrodidae, whitefly, *Glycine max*, IPM, pest monitoring.

2.2 Introduction

Soybean *Glycine max* (L) (Merr) is the most produced and consumed oilseed globally and the second good within the crop and livestock category traded on the planet (FAOSTAT, 2020; USDA, 2021). Soybean trade (exportation + importation) was over US$170 billion in 2020 (FAOSTAT, 2020). It is estimated that 383.59 million tons of soybean will be grown on 132.52 million hectares worldwide in the 2021/2022 harvest (USDA, 2021). Besides, this legume is relevant for animal and human nutrition, the biofuel industry, textiles, paints, and plastics (Chen et al., 2012). The attack of insect pests reduces soybean yield by 25%, and one of the threats is the whitefly *Bemisia tabaci* (Gennadius 1889) (Hemiptera: Aleyrodidae) (Gaur & Mogalapu, 2018).

The whitefly, *Bemisia tabaci* (Gennadius 1889) (Hemiptera: Aleyrodidae), is a devastating pest in tropical and subtropical agricultural regions worldwide (CABI, Li et al., 2021). Studies indicate that this pest reproduces in 36 genera, 76 families, and more than 1000 species of plants (Abd-rabou & Simmons, 2010; Willis, 2017). Whitefly is a primary pest to vegetables, grain legumes, ornamentals, and cotton production (Byrne & Bellows, 1991; Barro et al., 2011). This pest damages plants by feeding on phloem sap, injecting toxins, excreting the substrate for sooty mold growth, and transmitting over 100 plant viruses (Byrne, 1991; Jones 2003, Li et al., 2021). In addition, *B. tabaci* has evolved resistance to most insecticide classes, and resistance cases have been reported in 165 countries (Willis, 2017; Horowitz et al., 2020).

After establishing in Brazilian fields, *Bemisia tabaci* became a key pest in soybean culture (Lima et al., 2002; Pozebom et al., 2020). On average, just one whitefly per leaf reduces soybean yield by 31.24 kg. ha^{-1} (Padilha et al., 2021). Since the last decade, whitefly control faliture and outbreaks in soybean fields have intensified in Brazil (Dângelo et al., 2018, Arneman, 2019; Pozebom et al., 2020, Almeida et al 2021). In this condition, whitefly population would reduce soybean yield in 300 kg per ha (APRO-SOJA, 2022). To minimize the impacts of *Bemisia tabaci* on soybean crops, integrated pest management (IPM) programs must be developed and implemented.

Sampling plans are essential for IPM decision-making tools (Binns & Nyrop 1992, Arcanjo et al., 2021). A standardized sampling plan is the first step in tracking pest densities in an IPM program (Lopes et al., 2019; Cordeiro et al., 2021). This sampling plan must contain a sampling unit, a technique to scout the pest, and the number of samples needed to determine a pest density or damage, observing the data frequency distribution (Lopes et al., 2019; Silva et al., 2019). Also, it could provide scientific information to develop and validate sequential sampling plans and smartphone apps. A solid conventional sampling plan must combine precision with reasonable cost, time, and straightforward performance to increase farmers' usability (Pinto et al., 2017; Silva et al., 2019b; Cordeiro et al., 2021).

Methods to scout soybean fields in Brazilian soybean come from studies developed in North America, and they are not realistic for some pests in the Neotropical region (Boyer et al., 1969; Shepard, 1974; Panizzi 1997). In the case of whitefly, a lack of knowledge about sampling methods in tropical soybean areas persists. Czepack et al. (2018) proposed a sampling plan to scout whitefly nymphs by direct count on soybean leaflet subsections but did not investigate adults and other sampling techniques. Sampling adults would be more accessible to scout. Recently, Barros et al., 2021 studied a proximal sensing monitoring strategy for whiteflies in Brazilian soybean fields. However, a drone sensor mounted on the area did not validate the proposal. Besides, this method showed less accuracy (69.57%) in scouting lower whitefly density than in medium (85.71%). Finally, remote sensing techniques applying hyperspectral or multispectral cameras mounted on drones remain expensive in the neotropical region.

Despite these recent contributions, more studies are demanded to improve *B. tabaci* scouting in commercial soybean fields in the Neotropical areas. Hence, this study aimed to determine a novel standardized sampling plan for *Bemisia tabaci* adults throughout commercial soybean fields in the neotropical region, which can be adopted in all plant stages.

2.3 Material and Methods

2.3.1 Experiment settings

This study was performed in 26 commercial soybean fields at vegetative, flowering, and fruiting stages in Formoso do Araguaia (11°55'23.80"S; 49°41'40.70"O, 240 m altitude) and Gurupi (11°48'10.30"S; 49°00'29.30"O, 287 m altitude), state of the Tocantins, Brazil from 2017 to 2021. The soybean variety cultivated was the M 8808 IPRO (Intacta) with a determined cycle (Agro Bayer Brasil). Soybean crops were spaced at 0.45 m between rows, and the sowing was carried out at 0.02 m deep with 13 seeds per meter, resulting in 288,888.89 plants per crop. The assessed areas consisted of 7 to 25 ha.

In this research, whitefly adults were evaluated in regard to establishing a standardized sampling plan following five steps: 1- sample unit determination, 2- Sample technique evaluation, 3- Frequency distribution of whitefly density, 4- Optimum number of samples calculation, and 5- Costs and time of sampling process.

2.3.2 Sample unit determination

This part of the research was carried out in three soybean fields at distinct phenological stages, vegetative (V4), flowering (R2), and fruiting (R5). In this step, a site on soybean plants to observe whitefly adults throughout the phenological stages has been selected (sample unit). Regarding that, soybean leaves were enumerated top-down in each plant evaluated. After that, whitefly's adults in the entire plant canopy were assessed in 40 plants per field by a direct count in each soybean leaf. Sample unit selection was performed according to leaf frequency occurrence, precision, representativeness, and practicability criteria (Podoler and Rogers 1975; Southwood, 1978; Lopes et al., 2019). First, considering the occurrence criteria, leaves with a presence frequency below 80% were eliminated because they would be laborious to assess (Lopes et al., 2020). Second, according to the precision criteria, samples with *B. tabaci* density with 25% up of relative variance (R.V) were removed to avoid unfeasible sampling plans (Southwood, 1978; Gonring et al., 2020). R.V of *Bemisia tabaci* adults for each sample tested was calculated following this formula 1:

$RV = 100\, x\, (SE/\, \bar{x})$ (1)

Where SE is the standard error and $\bar{x}$ means insect average.

Subsequently, through the representativeness criteria, whitefly density in the whole plant (absolute count) was submitted to Pearson correlation with each sample unit candidate (relative count), and samples presenting non-significative correlation at $p < 0.05$ were disqualified (Silva Jr et al., 2020). Finally, the sample unit designated to scout whitefly adults in tropical soybean fields was the one that matched these assumptions above in all phenological stages tested following the practicability requirement (Araújo et al., 2019).

2.3.3 Sample technique evaluation

A sample technique study to assess *B. tabaci* adults in soybean plants was performed in three fields at distinct stages: vegetative, flowering, and fruiting. In each area, whitefly density was determined in 20 plants randomly by direct counting, plastic tray, and beating cloth approaches. These techniques were tested because they have been adopted to scout sucking insects in soybean operations. Besides, the time required to scout the pest was recorded for every method. Samplings were taken in the apical section of the soybean canopy. The apical part of the soybean canopy presents most whitefly adults. Besides, it has been selected for monitoring whitefly and other sucking insects in soybean and other row crops (Macêdo et al., 2018; Silva et al., 2019; Shah et al., 2020). The sampling technique was selected according to the precision and speed criteria. A sampling method must present an R.V less than 25% whitefly density and a loweest evaluation time to be elected (Gonring et al., 2020; Carmo et al., 2021).

Insect density captured in the fields by each sampling method and the time spent to evaluate them were compared by ANOVA GLM followed by Tukey test post hoc ($p <$ 0.05) employing R software packages MASS and Multcomp, respectively (Ripley et al., 2013; Hothorn et al., 2016; R Core Team, 2019). For counting data, Poisson or Quasipoisson (controlling overdispersion) distribution families were employed in GLM models, while for continuous data (time to evaluate the insects), Gamma distribution was adopted to build GLM models (Helmreich, 2015).

2.3.4 Frequency distribution of whitefly density

The frequency distribution of whitefly density was accessed upon a total of 21 soybean fields (5, 6, and 10 fields at vegetative, flowering, and ten fruiting stages, respectively). About 200 plants per field were examined upon a plastic beating tray technique against soybean canopy's apex (previously selected method). Whitefly adults' densities were tested against the Negative binomial, Positive binomial, and Poisson frequency distribution throughout a chi-squared test (Lopes et al., 2019; Mota et al., 2022). Pest density has fitted on a frequency distribution when expected and observed chi-square values were non-significative at $\alpha = 0.05$. A frequency distribution that fitted at least 70% of the soybean fields was adopted for selecting a formula to calculate the number of samples required for scouting *B. tabaci* in these tropical soybean fields (Young & Young, 1998; Lopes et al., 2019; Santana et al., 2021).

2.3.5 Optimum number of samples calculation

First, the aggregated parameters values (k) of the Negative binomial frequency distribution were calculated for each field following formula 2 as *Bemisia tabaci* density data fitted on this distribution on most of the soybean fields evaluated.

$K = X^{-2} / (S^2 - \bar{x})$, (2)

Where K is the aggregated parameter of the Negative binomial frequency distribution, x is the average whitefly sampled per field, and S is the variance of this pest sampling data.

After that, a K common (Kc), which is an aggregated parameter of the Negative binomial distribution for all the evaluated soybean fields, was employed to calculate the ideal number of samples to scout *Bemisia tabaci* in soybean fields in consonance with formula 3:

$NA = \frac{1}{C^2} x \left(\frac{1}{\bar{x}} + \frac{1C}{Kc}\right)$, (3)

Where NA is the ideal number of samples to scout *Bemisia tabaci* in soybean fields, C is the maximum sampling error (25%) allowed in the sampling plan, and Kc is the K common parameter for all soybean fields in this study (0.8145).

Error-values used to calculate the number of samples varied from 1 to 25%, being 20% the error selected to determine the number of samples in this standardized sampling plan (Southwood, 1978).

2.3.6 Sampling time and costs

This part of the research was conducted for five soybean-sized scenery: 10, 20, 30, 50, and 100 ha. These areas were selected because they represent general soybean operations in the Neotropical lands. The time and costs spent in the sampling process were determined for each situation.

The sampling time consists of the time to scout the pest and walk between samples. Scouting time is the time to assess each sampling unit, beat the plants against the plastic tray, and count the insects. The time spent walking between the samples (49 plants, previously determined) is the second part of the sampling time. The scouting and movement time spent between sampling sites were measured in the field.

Sampling costs comprised scouting time, material (pencil, rubber, paper, and clipboard), and labor costs. Labor cost was calculated for Brazil and the United States as both are the major world soybean producers, representing 70% of the global soybean production (USDA, 2022). For the U.S., a wage average of the soybean grower states was adopted (U.S Department of Labor).

Then, the cost for sampling *B. tabaci* in soybean fields was calculated through the formula 4:

$$SC = MC + ((ST + WT)\ X\ \ LC)\ , (4)$$

Where SC is the sampling cost to evaluate *Bemisia tabaci* in soybean fields, MC is the material cost, considering one year of durability, S.T. is the scouting time of *B. tabaci*, W.T. is the time spent moving between samples, and L.C. is the labor costs (worker wage plus benefits).

2.4 Results

2.4.1 Sample unit

The apex leaves of the plants present a fairly occurrence frequency rate (at least 80%) in the legume canopy across the vegetative, flowering, and fruiting stages. Then, leaves localized on the top-down positions 1st to 6th; 1st to 14; and 1st to 30 at the vegetative, flowering, and fruiting phases, respectively, are potential candidates to be a sample unit (Fig. 1a, 1b, and 1c). However, the relative variance was superior to 25% in most of these sample unit candidates at the flowering and fruiting stages. After this criteria, the remaining sampling candidates were 1st to 6th; 1st to 3rd; and 1st to 4th, 6th, and 7th at the vegetative, flowering, and fruiting stages (Fig. 1a, 1b, and 1c). Finally, Pearson correlation analyses between whitefly relative and absolute density were significant ($p<0.05$) to the leave positions 1st to 3rd; 1st, 3rd; and 1st to 3rd at vegetative, flowering, and reproductive stages. As the most apical leaf presents a positive and significant correlation in all the phenological phases, the first leaf at the top of the plant canopy was elected as an ideal sampling plan (Fig. 1a, 1b, and 1c).

2.4.2 Sample technique

In the vegetative stage, the direct count technique trapped more whitefly adults, followed by beating on a plastic tray and beating cloth methods (GLM $_{\text{Poisson}}$; $\chi^2 = 55.756$, df = 57, $p < 0{,}0001$) (Fig. 2a). In the flowering phase, the beating on the plastic tray captured more insects than the other techniques ($F_{2,57} = 10.655$, $p < 0{,}001$). Conversely, no differences were noticed among the sampling techniques at the fruiting stage (GLM $_{\text{Quasipoisson}}$; $F_{2,57} = 2.722$, $p = 0.07$) (Fig. 2a).

In the vegetative stage, the whiteflies sampled with the beating clothing yielded relative variance above the upper limit of 25%. Although, for all the other techniques and phenological soybean phase, relative variance was lower than 25%. As the beating cloth technique has not presented a relative variance inferior to 25% in all plant phenological stages, it would not be a practical sampling technique (Fig. 2b). Finally, the sampling timing was different according to plant stage and sampling technique. In the vegetative stage, the direct count technique has consumed more time than other methods, followed by beating cloth and plastic tray approaches (GLM $_{\text{Gamma}}$; $\chi^2 = 23.043$, df = 27, $p < 0{,}0001$). In the flowering (GLM $_{\text{Gamma}}$; $\chi^2 = 3.464$, df = 27, $p < 0{,}0001$) and fruiting (GLM $_{\text{Gamma}}$; $\chi^2 = 25.218$, df = 27, $p < 0{,}0001$) stages, direct count and beating clothe methods spent more time sampling than the plastic tray technique (Fig. 2c). Therefore, beating a plastic tray is the most feasible technique.

2.4.3 Number of Samples

Fields data have adjusted in the Negative Binomial frequency distribution in 100% of the vegetative and flowering stages. In the reproductive stage, the Negative Binomial frequency distribution has fitted in 9 out of 11 field data (81,82%) (table 1; Fig. 4). The general fit of this distribution was 90% (18 out 20 of field data) (table 1; Fig. 4). However, field data have adjusted poorly against Poisson and Positive Binomial frequency distribution. Poisson distribution has fitted only in 2 out of 20 field data (10%), and these two adjusted fields occurred just in the vegetative stage. Nevertheless, the Positive Binomial frequency distribution has not fitted in any field data (table 1; Fig.4). Finally, the common aggregated parameter of the Negative Binomial frequency distribution on the evaluated fields is 0.8145 with significative slope ($F_{1,17}$ = 44.93, $p<0.05$) and non-significative intercept ($F_{1,17}$ = 0.26, $p<0.05$).

The number of samples to be evaluated per field admitting errors (5, 10, 15, 20, and 25) were 784, 242, 87, 49, and 31. Then 49 samples at a 20% of error (upon this error number of samples do not reduce substantially) were elected as the ideal number of samples to scout *B. tabaci* adults upon a plastic tray beating technique at the vegetative, flowering, and fruiting stages in soybean crops following this conventional sampling plan guidance (Fig. 5).

2.4.4 Timing of sampling and Costs

In fields of 10, 20, 30, 50, and 100 ha, at distances of 3804.22; 5379.98; 6589.10; 8506.49; and 120030.00 meters between each sampling site, the estimated time to walk among samples would be 39, 55, 67, 86, and 122 minutes respectively. The time to access and scout the pest by the plastic tray technique would be 10 minutes for all field sizes. Hence, the total time to scout *B. tabaci* in soybean fields of 10, 20, 30, 50, and 100 ha is 49, 65, 77, 96, and 132 minutes. The walking movement among sites represents 79.59, 84.62, 87.01, 89.58, and 92.42 percent of the total time to monitor *B. tabaci* adults in fields 10, 20, 30, 50, and 100 ha, respectively. Then, the cost to scout *B. tabaci* in Brazilian crops is 3.93, 5.10, 5.99, 7.38, and 10.03 dollars, while in the United States of America (U.S.) is 7.42, 9.84, 11.65, 14.53, and 19.98 in fields of 10, 20, 30, 50, and 100 ha, respectively. Finally, the sampling cost per hectare would be 0.39, 0.26, 0.20, 0.15, and 0.10 dollars in Brazilian fields and 0.74, 0.49, 0.39, 0.29, 0.20 dollars in the US.

2.5 Discussion

This research provides a solid protocol for scouting whitefly *Bemisia tabaci* in soybean crops. In this standardized sampling plan, whitefly adults must be examined in the apical part of the plant canopy through a plastic beating tray technique in 49 plants per field across the soybean phenological stage (vegetative, flowering, and fruiting).

The affordable sampling unit for monitoring whitefly in soybean crops is the apical part of the plant. The apical part of the plant hosts the most considerable amount of the whitefly population; it also meets the criteria of sample frequency (> 80%), precision (relative variance < 25%), and representativity (positive and significative correlation between the sampling unit and absolute density). Besides, the same site could be accessed throughout the soybean stages, making the monitoring process practical. Finally, it would facilitate the sampling adoption at a field level (Mota et al., 2022; Silva et al., 2020b).

The ideal technique to scout whitefly adults determined in our investigation is beating the plant's apex against a plastic tray. This technique makes it possible to assess whitefly adults with precision and the lowest time. Although direct count attends the precision criteria (relative variance <25%), it expends more time than the plastic tray technique on all plant phenological scenery. Direct counting is laborious and frequently spends extra time than other methods. The beating cloth technique has been widely adopted to monitor soybean pests in Brazil and the United States of America. Although, in our study, this sampling method was not appropriate, especially in the vegetative stage, yielding a relative variance of 68.82%. Also, this method produces more time than the beating tray technique as the processes required to set up the beating cloth between the rows, shake the plants, and count the insects are more laborious than beating the plants against the plastic tray. Therefore, the plastic tray method is the most appropriate for evaluating whitefly adults in soybean crops (Pinto et al., 2017; Lopes et al., 2019).

The whitefly density data fit on the negative binomial frequency distribution model. This frequency distribution generally is suitable for counting data with a variance larger than the average (Helmreich, 2015). Also, it is employed in many sampling plans and yields an affordable number of samples per plot. The existence of an aggregated K common for all crops in this study indicates that this plan can be applied in others soybean fields in tropical conditions. This protocol suggests that 49 plants should be evaluated per field, considering an error of 20%. Generally, 25% of error is the maximum admitted in a conventional sampling plan. Although, 20% will enhance the sampling plan precision without substantially boosting sampling costs (Bliss & Owen 1958; Lopes et al., 2019).

The sampling time is strongly influenced by walking between sample time and the plot size to scout the pest. In most cases, up 80% of the time is spent walking, reaching 92,42% in plots with 100 ha. Nevertheless, the scouting time for evaluating all the 49 samplings is rapid (10 minutes). The cost of the sample is firmly shaped by the plot size and country (Brazil or the United States). In Brazil, the sampling cost of evaluating a small plot (10 ha) is US$ 3.93, while in a larger field, the cost is US$ 10.03. However, in the United States, the sampling cost is almost twice the Brazilian cost. This fact is explained by the labor cost, approximately 4.5 times more than the Brazilian agriculture labor price (U.S Department of Labor; Governo Brasileiro, 2022).

Digital agriculture based on artificial intelligence tools and remote sensing is the trend of monitoring insect pest. Although, these technologies must not fit in some systems as sucking pests because it is not simple to determine pest injury damage-relationship by reflectance, mainly when a pest attack is beginning. Also, monitoring insects with digital images through Convolution Neural Network technology without batch pheromones would be another challenge, as identifying and counting insects in traps with a range of non-target organism and impurities (dusk, leaf, e.g.) difficult pest tracking. Therefore, classic studies would fit in digital agriculture when deployed in smartphone apps (Barbedo, 2020; Rustia et al., 2020; Barros, et al., 2021).

This sampling protocol is the base for developing other studies about monitoring systems on whitefly soybean. The next step of this research would be establishing a decision-making plan based on sequential sampling plans. Also, this study would be implemented on smartphone devices which growers would input personal details about pest control.

2.6 Conclusion

This study provides novel guidance to scout *Bemisia tabaci* adults in Neotropical soybean crops through a standardized sampling plan. This sampling plan consists of an ideal sampling unit, an apical part of the soybean canopy. Also, the most appropriate sampling technique is beating the apical site of the plant against a plastic tray. Forty-nine soybean plants must be assessed in soybean fields in this feasible conventional sampling plan. The timing and costs of this sampling method are rapid (maximum of 5 min per ha) and affordable (the ultimate cost of US$ 0.74 per ha) to meet the soybean operations requirement. Hence, this standardized sampling plan is precise, representative, fast, and presents a reasonable cost. Finally, it is the first step to determining sequential sampling plans and could be implemented in digital agriculture devices.

2.7 Acknowledgment

Financial support was provided by the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq); Coordination of Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Finance Code 001); and Minas Gerais State Foundation for Research Aid (Fundação de Amparo a Pesquisa do Estado de Minas Gerais– FAPEMIG).

2.8 References

Abd-Rabou, S., & Simmons, A. M. (2010). Survey of reproductive host plants of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Egypt, including new host records. *Entomological News*, *121*(5), 456–465.

Agrobayer Brasil. https://www.agro.bayer.com.br/essenciais-do-campo/sementes/monsoy/m-8808-ipro. Accessed in June 2022.

Almeida, M.F., Tavares, C.S., Araújo, E.O., Picanço, M.C., Oliveira, E.E., & Pereira, E.J. G. (2021). Plant Resistance in Some Modern Soybean Varieties May Favor Population Growth and Modify the Stylet Penetration of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Journal of Economic Entomology*, *114*(2), 970–978.

APROSOJA. Infestação de mosca-branca causa perda de até cinco sacas por hectare em Mato Grosso na safra 16/17. https://www.semagro.ms.gov.br/infestacao-de-mosca-branca-causa-perdas-de-ate-cinco-sacas-de-soja-por-hectare-em-mato-grosso-na-safra-1617/. Acessed on June, 2022.

Araújo, T.A., Araújo, L.H.A., Silva, N.R., Luz, C.E.A., da Silva, É.M., Moreira, M.D., ... & Bastos, C.S. (2019). Standardized sampling plan for *Aphis gossypii* based on the cotton cultivar, plant phenology and crop size. *Journal of Applied Entomology*, *143*(8), 893–901.

Arcanjo, L.P., Silva, E.M., Araújo, T.A., Crespo, A.L.B., Júnior, P.A.S., Gomes, G.B.O., & Picanço, M.C. (2021). Decision-making systems for management of the invasive pest *Neoleucinodes elegantalis* (Guenée) (Lepidoptera: Crambidae) in commercial tomato crops according to insecticide spray method and plant stage. *Crop Protection*, *140*, 105408.

Arnemann, J.A., Pozebon, H., Marques, R.P., Ferreira, D.R., Patias, L.S., Bevilaqua, J. G., ... & Ferreira, P. E. (2019). Managing whitefly on soybean. The Journal of Agriculture Science. *11*(9).

Barbedo, J.G.A. (2020). Detecting and Classifying Pests in Crops Using Proximal Images and Machine Learning: A Review. *AI*, 1(2), 312–328

Barro, P. J., Liu, S. S., Boykin, L. M., & Dinsdale, A. B. (2011). *Bemisia tabaci*: a statement of species status. *Annual review of entomology*, *56*(1), 1–19.

Barros, P. P., Schutze, I. X., Iost Filho, F. H., Yamamoto, P. T., Fiorio, P. R., & Demattê, J. A. (2021). Monitoring *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) infestation in Soybean by proximal sensing. *Insects*, *12*(1), 47.

Binns, M.R., & Nyrop, J. P. (1992). Sampling insect populations for the purpose of IPM decision making. *Annual review of entomology*, *37*(1), 427–453.

Boyer, W.P; Dumas, W.A. Plant-shaking methods for soybean insect survey in Arkansas. In Survey methods for some Economic insects; USDA: Forrest City, AR, USA, 1969; pp. 92–94.

Byrne, D.N., & Bellows, T.S. (1991). Whitefly biology. *Annual review of entomology*, *36*, 431–457.

CABI—Center for Agriculture and Biosciences International. Invasive Species Compendium. *Bemisia tabaci* (Tobacco Whitefly). 2019. Available online: http://www.cabi.org/isc/datasheet/8927 (accessed on 13 September 2020).

Carmo, D.D.G., Lopes, M.C., Araújo, T.A., Ramos, R.S., Soares, J.R.S., Silva Paes, J., & Picanço, M.C. (2021). Conventional sampling plan for green peach aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae), in bell pepper crops. *Crop Protection*, *145*, 105645.

Chen, K.I., Erh, M.H., Su, N.W., Liu, W.H., Chou, C.C., & Cheng, K.C. (2012). Soyfoods and soybean products: from traditional use to modern applications. *Applied microbiology and biotechnology*, *96*(1), 9–22.

Computing, Vienna, Austria. URL https://www.R-project.org/.

Czepak, C., Coelho, A.S.G., Rezende, J.M., Le Senechal Nunes, M., Weber, I.D., Silvério, R.F., & Albernaz-Godinho, K.C. (2018). *Bemisia tabaci* MEAM 1 population surveys in soybean cultivation. *Entomologia Experimentalis et Applicata*, *166*(3), 215–223.

Dângelo, R.A.C., Michereff-Filho, M., Campos, M.R., Silva, P.S., & Guedes, R.N.C. (2018). Insecticide resistance and control failure likelihood of the whitefly *Bemisia tabaci* (MEAM1; B biotype): a Neotropical scenario. *Annals of Applied Biology*, *172*(1), 88–99.

FAOSTAT. Crops and livestock products. 2020. https://www.fao.org/faostat/en/#data/QCL. Accessed in June 2022.

Gaur, N., & Mogalapu, S. (2018). Pests of soybean. In *Pests and their management* (pp. 137–162). Springer, Singapore.

Gonring, A.H., Walerius, A.H., Picanço, M.M., Bacci, L., Martins, J. C., & Picanço, M.C. (2020). Feasible sampling plan for *Tuta absoluta* egg densities evaluation in commercial field tomato. *Crop Protection*, *136*, 105239.

Google shoping. Multispectral camera. shorturl.at/hotAM. Acessed in June 2022.

Governo Brasileiro. Medida provisória nº 1.091 de 30 de dezembro de 2021. https://legislacao.presidencia.gov.br/atos/?tipo=MPV&numero=1091&ano=2021&ato=dc2MzYU9UMZpWTc91. Acessed in June 2022.

Helmreich, J.E. (2015). Statistics: An introduction using R. *Journal of Statistical Software*, *67*, 1–4.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A., Scheibe, S., & Hothorn, M.T. (2016). Package ‘multcomp’. *Simultaneous inference in general parametric models. Project for Statistical Computing, Vienna, Austria*.

Jones, D.R. (2003). Plant viruses transmitted by whiteflies. *European journal of plant pathology*, *109*(3), 195–219.

Li, Y., Mbata, G.N., Punnuri, S., Simmons, A. M., & Shapiro-Ilan, D.I. (2021). *Bemisia tabaci* on vegetables in the Southern United States: incidence, impact, and management. *Insects*, *12*(3), 198.

Lima, L.H.C., Campos, L., Moretzsohn, M.C., Návia, D., & De Oliveira, M.R.V. (2002). Genetic diversity of *Bemisia tabaci* (Genn.) populations in Brazil revealed by RAPD markers. *Genetics and Molecular Biology*, *25*, 217–223.

Lopes, M.C., Ribeiro, A.V., Costa, T.L., Arcanjo, L.P., Farias, E.S., Santos, A.A., ... & Picanço, M.C. (2019). Practical sampling plan for *Liriomyza huidobrensis* (Diptera: Agromyzidae) in tomato crops. *Journal of Economic Entomology*, *112*(4), 1946–1952.

Macêdo, R.V.B.T., Sarmento, R.A., Pereira, P.S., Lima, C.H.O., Deus, T.L.L.B., Ribeiro, A.V., & Picanço, M.C. (2019). Sampling plan for *Bemisia tabaci* (Hemiptera: Aleyrodidae) in melon crops. *Florida Entomologist*, *102*(1), 16–23.

Padilha, G., Pozebon, H., Patias, L.S., Ferreira, D.R., Castilhos, L.B., Forgiarini, S.E., ... & Arnemann, J.A. (2021). Damage assessment of *Bemisia tabaci* and economic injury level on soybean. *Crop Protection*, *143*, 105542.

Panizzi, A.R.; Corrêa, B.S; Grazzoni, D.L; De Oliveira, E.B.; Newman, G.G.; Turnipseed, S.G. Insetos da Sona no Brasil; Embrapa-CNPSo: Londrina, Brazil, 1997; p.20, (Embrapa-CNPSo. Boletim Técnico,1).

Pinto, C.B., Sarmento, R.A, Galdino, T.V.S, Pereira, P.S., Barbosa, B.G., Lima, C.H.O, ... & Picanço, M.C (2017). Standardized sampling plan for the thrips *Frankliniella schultzei* (Thysanoptera: Thripidae) on watermelon crops. *Journal of Economic Entomology*, *110*(2), 748–754.

Podoler, H., & D. Rogers. 1975. A new method for the identification of key factors from life-table data. *Jounal of Animal Ecology*. 44: 85–114.

Pozebon, H., Marques, R.P., Padilha, G., O´ Neal, M., Valmorbida, I., Bevilaqua, J. G., ... & Arnemann, J.A. (2020). Arthropod invasions versus soybean production in Brazil: a review. *Journal of economic entomology*, *113*(4), 1591–1608.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M.B. (2013). Package ‘mass’. *Cran r*, *538*, 113–120.

Santana, A.S, Lima, A.P.S., Santana, E.D.R., Santos, A.C.C., Cristaldo, P.F., Araújo, A. P.A., & Bacci, L. (2021). Development of conventional sampling plans for egg masses and nymphs of citrus blackfly *Aleurocanthus woglumi* Ashby (Hemiptera: Aleyrodidae). *Crop Protection*, *149*, 105777.

Santos, R.C., Lopes, M.C., Sarmento, R.A., Pereira, P.S., Picanço, M.M., Santos, W.P., ... & Picanço, M.C. (2021). Conventional sampling plan for thrips in tropical soybean fields. *Crop Protection*, *148*, 105740.

Shah, M.A., Sridhar, J., & Sharma, S. (2020). Studies on the within-plant distribution of whitefly, *Bemisia tabaci* (Gennadius) in potato and development of a sequential sampling plan. International Journal of Pest Management, 1–10.

Shepard, B. M; Carner, G.R.; Turnipseed, S.G. A Comparison of Three Sampling methods for arthropods in soybean. Environ. Entomolo. 1974, 3, 227–232.

Silva Jr, F.S., Lopes, M.C., Farias, E.S., Sarmento, R.A., Pereira, P.S., & Picanço, M.C. (2020). Standardized sampling plan for common blossom thrips management in melon fields from North Brazil. *Crop Protection*, *134*, 105179.

Silva, E.M.D.P., Araújo, T.A.D., Ramos, R.S., Arcanjo, L.P., Carmo, D.D.G.D., Cavalleri, A., & Picanço, M. C. (2019). Conventional sampling plan for common blossom Thrips, *Frankliniella schultzei* (Thysanoptera: Thripidae), in bell pepper. *Journal of economic entomology*, *112*(3), 1447–1453.

Silva, É.M., Bacci, L., Picanço, M.C., Gonring, A.H.R., Neves, L.G., Silva, R.S., & Martins, J.C. (2020b). Sampling plan of *Tetranychus mexicanus* on passion fruit vines. *International Journal of Pest Management*, *67*(4), 269–278.

Silva, É. M., Crespo, A. L., Farias, E. S., Bacci, L., Queiroz, R. B., & Picanço, M. C. (2019b). Conventional sampling plan for scouting *Neoleucinodes elegantalis* (Lepidoptera: Crambidae) eggs on tomato fruits. *Journal of Economic Entomology*, *112*(5), 2433–2440.

Southwood, T.R.E. 1978. Ecological methods: with particular reference to study of insect populations. Chapman and Hall, New York.

Souza, M.D.S.C., Lopes, M.C., Arcanjo, L.P., Mesquita, A.L.M., Pereira, R.D.C.A., Silva, R.S., & Picanço, M.C. (2022). Designing a robust strategy to scout a key pest in cashew orchards. *Crop Protection*, *157*, 105954.

U.S Department of Agriculture. USDA. Foreign Agricultural Service. 2022. https://www.fas.usda.gov/. Accessed in June, 2022.

U.S Department of Labor. https://www.dol.gov/agencies/whd/minimum-wage/state. Accessed in 2022.

Willis, K. (2017). *State of the world's plants 2017*. Royal Botanics Gardens Kew.

Young, L.J., and J.H. Young. 1998. Statistical ecology: a population perspective. Kluwer Academic Publishers, Boston, MA.

2.9 Figures and tables

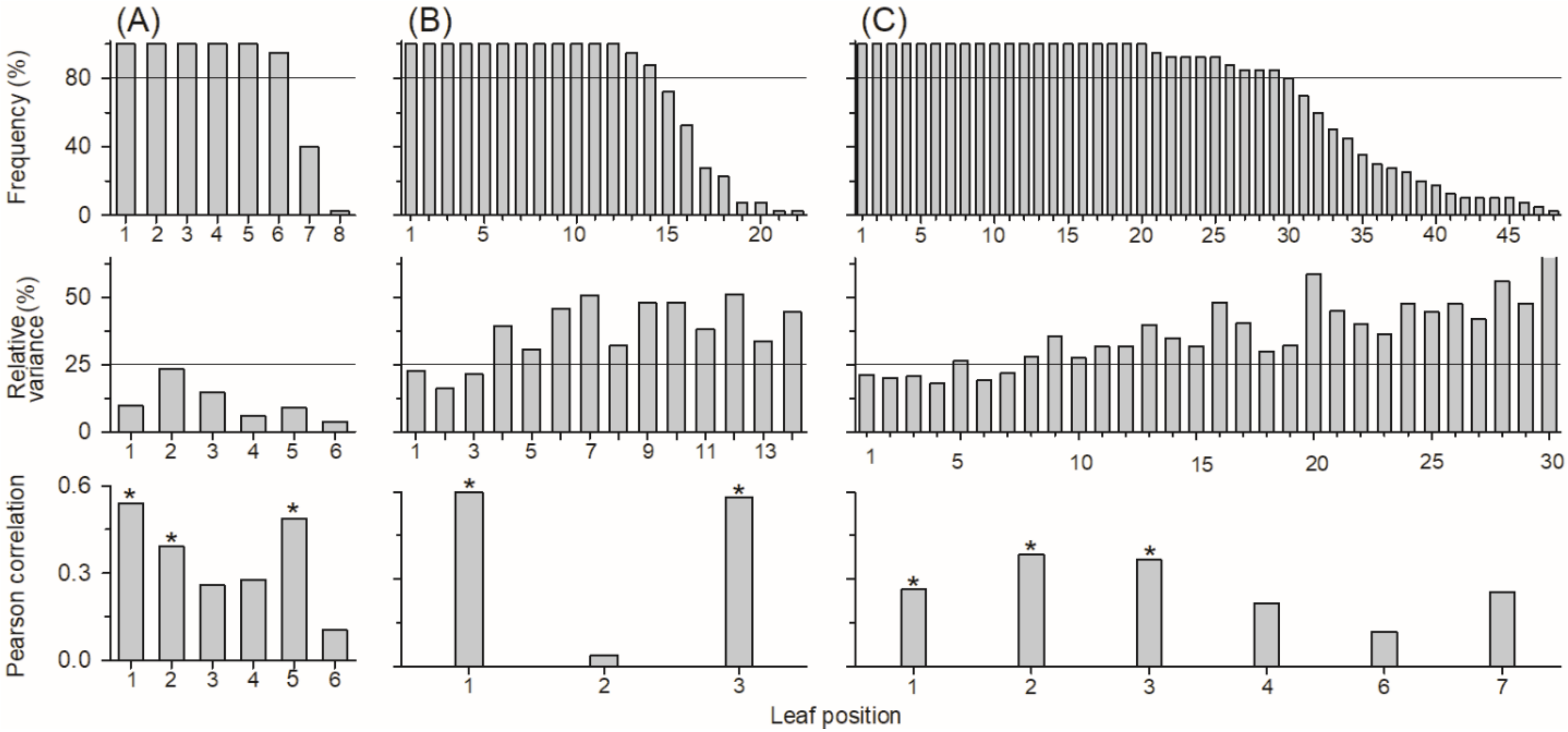

Figura 1. Sampling unit selection to evaluate *B. tabaci* adult population in soybean fields with plants at (A) vegetative, (B) flowering, and (C) fruiting stages: frequency of leaf occurrence in this position of the canopy, relative variance, Pearson correlation (r) between relative densities (adults. leaf^{-1}) with absolute density (adults. plant^{-1}). * Significant corrections and slopes according to the t-test at $p < 0.05$. Leaves numbered 1, 2, and n = 1st, 2nd, and nth leaf from the apex of the plant, respectively.

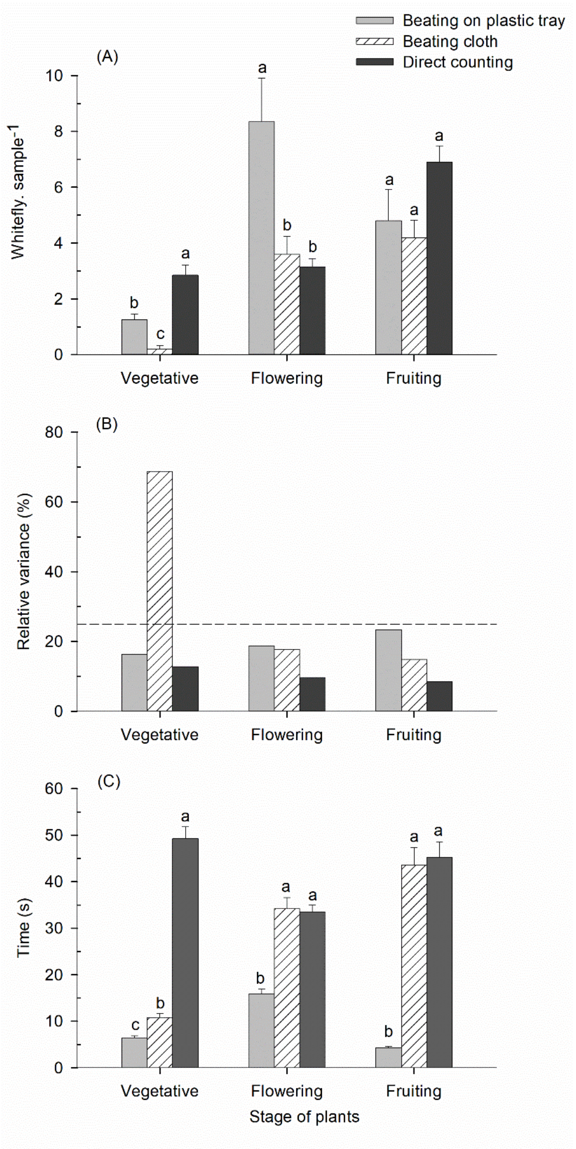

Figure 2. Sampling technique selection for scouting *Bemisia tabaci* adult's population in soybean fields at vegetative, flowering, and fruiting stages. (A) whitefly per sample (mean ± S.E.), (B) relative variance among densities (%), and (C) evaluation time (s) according to distinct sampling methods (Beating on a plastic tray, Beating cloth, and Direct counting). Different letters mean significative difference by Tukey's test ($p < 0.05$).

Table 1

B. tabaci adult's density per sample obtained through the beating on plastic tray technique in 20 soybean fields and frequency distributions fitting by χ2 test.

Field	Density (mean ± se)	Negative Binomial		Poisson		Positive Binomial	
		χ2	df	χ2	df	χ2	df
Vegetative stage							
1	0.55 ± 0.07	3.85^{ns}	2	37.12*	3	4.92E+02*	3
2	1.10 ± 0.11	2.66^{ns}	6	490.77*	7	3.21E+04*	7
3	1.45 ± 0.14	5.06^{ns}	7	239.64*	8	1.07E+06*	8
4	0.08 ± 0.02	0.01^{ns}	1	0.16^{ns}	1	2.45E+01*	1
5	0.04 ± 0.02	0.14^{ns}	1	3.83^{ns}	1	1.16E+01*	1
Flowering stage							
6	0.52 ± 0.07	1.10^{ns}	3	60.22*	4	7.38E+02*	4
7	4.29 ± 0.39	24.44^{ns}	15	3721.84*	16	1.40E+23*	16
8	0.85 ± 0.08	1.15^{ns}	4	21.84*	5	1.27E+04*	5
9	0.22 ± 0.03	1.84^{ns}	1	5.00*	1	6.09E+01*	1
Reproductive stage							
10	4.65 ± 0.34	19.03*	14	818.52*	15	2.52E+27*	15
11	1.42 ± 0.18	29.51*	7	321.69*	8	4.96E+10*	8
12	1.33 ± 0.14	26.71^{ns}	5	185.21*	6	6.04E+04*	6
13	0.16 ± 0.04	1.61^{ns}	1	7.00*	2	1.02E+02*	2
14	0.94 ± 0.08	5.42^{ns}	4	23.93*	5	1.31E+04*	5
15	1.11 ± 0.11	7.62^{ns}	4	39.96*	5	5.64E+05*	5
16	4.24 ± 0.23	18.05^{ns}	11	205.26*	12	7.05E+19*	12
17	1.54 ± 0.19	14.66^{ns}	9	1412.34*	10	5.32E+08*	10
18	1.11 ± 0.10	9.12^{ns}	5	160.38*	6	2.12E+04*	6
19	1.38 ± 0.13	12.02^{ns}	6	63.98*	7	8.01E+06*	7
20	0.35 ± 0.07	3.43^{ns}	3	67.93*	4	1.08E+03*	4

^{Ns} Non-significative. *Significant at 5% probability level according to the *F* test. df = degrees of freedom.

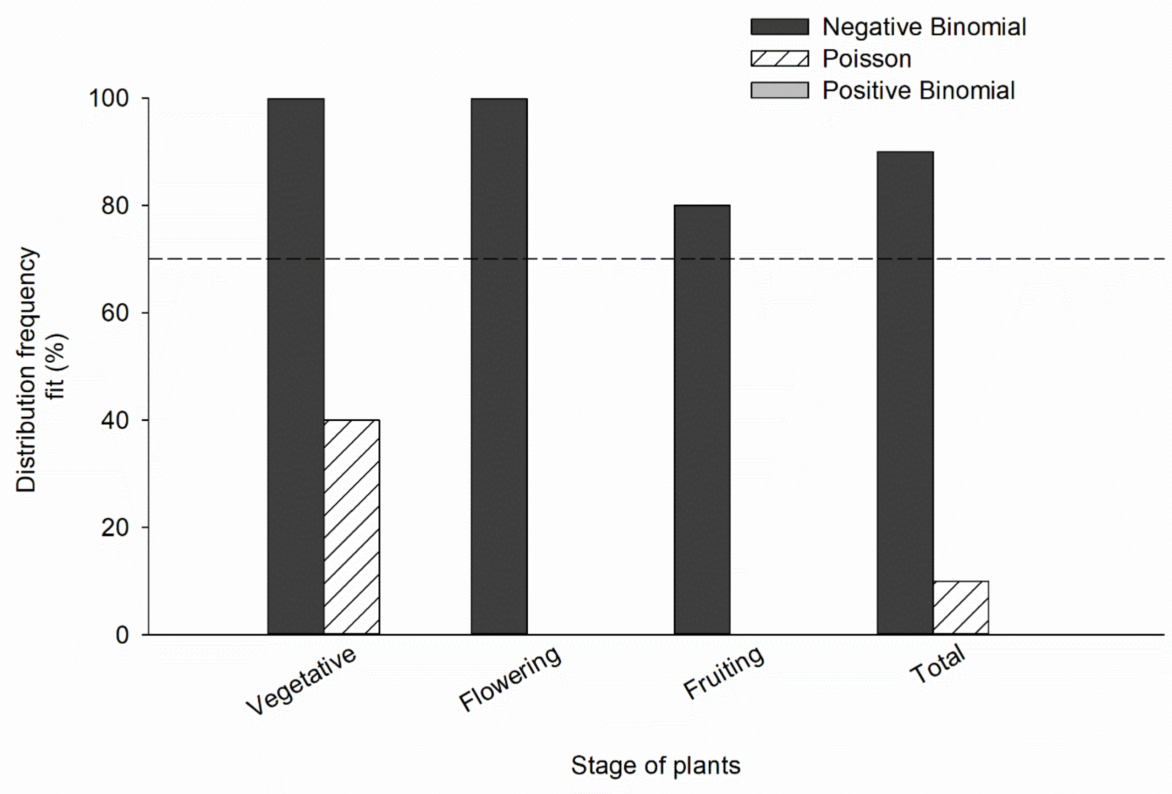

Figure 3. Distribution frequency fit (%) to the models (Negative Binomial, Poisson, and Positive Binomial) on *Bemisia tabaci* adults density sampled in 20 soybean fields at vegetative, flowering, and reproductive stages. Positive Binomial adherence was zero.

Table 2

ANOVA of *B. tabaci* densities (adults per sample) through plastic tray scout technique in 20 soybean fields, performed to evaluate the existence of a common aggregation parameter (kc) in a negative binomial distribution.

Variance source	df	Sum of squares	Mean squares	F
Slope 1/Kc	1	277.87	277.87	44.93*
Intercept	1	1.60	1.60	0.26^{ns}
Error	17	105.13	6.18	
Kc = 0.8145				

*Significant at the 5% probability level. ^{ns}Non-significant. df = degrees of freedom.

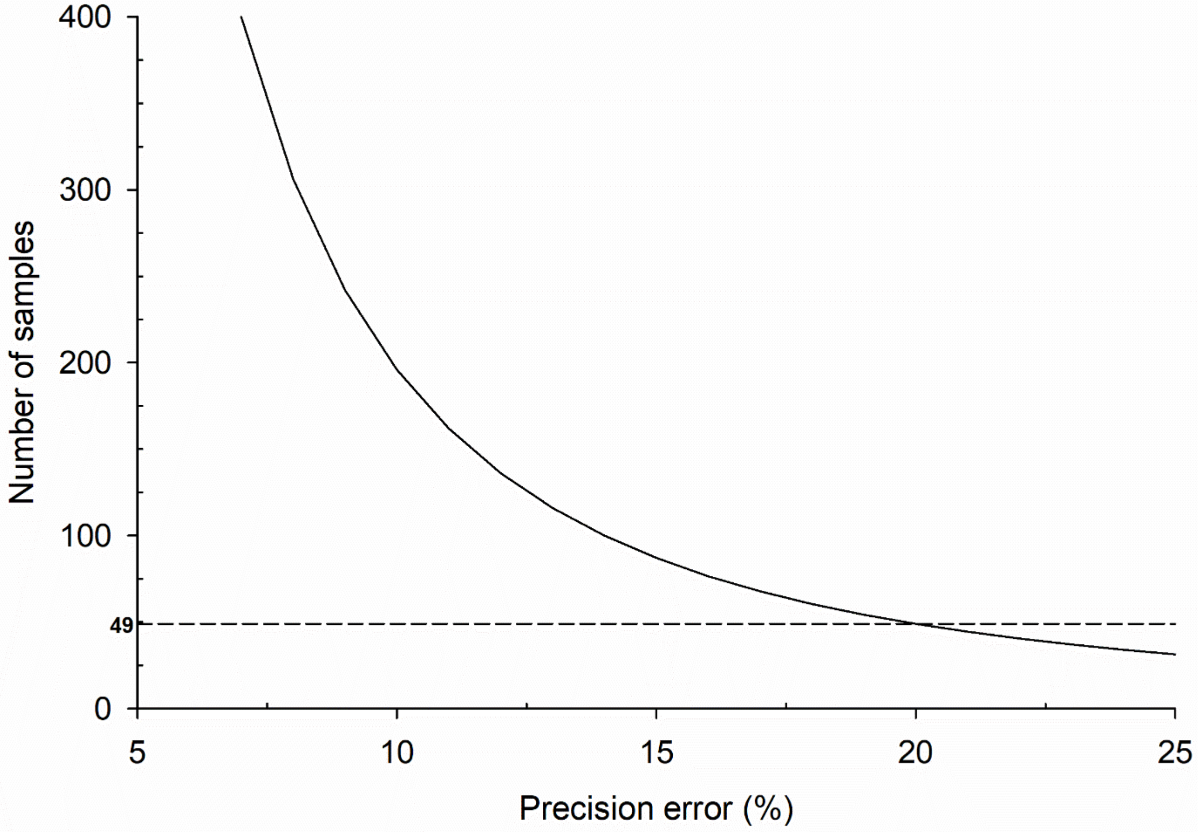

Figure 4. Number of samples required to scout *B. tabaci* adult population in soybean crops according to precision error levels by beating on tray technique.

Table 3

Walking distance, time, and cost in Brazil and United States required for sampling adults of whitefly *Bemisia tabaci* in soybean fields according to field size.

Field size (ha)	Distance walking (m)	Walking between samples		Sample evaluation		Total sampling time	Sampling cost* (US$)			
							Brazil		USA	
		Time	(%)	Time	(%)		per field	per hectare	per field	per hectare
10	3804.22	39 min	79.59	10 min	20.57	49 min	3.93	0.39	7.42	0.74
20	5379.98	55 min	84.62	10 min	15.48	65 min	5.10	0.26	9.84	0.49
30	6589.10	67 min	87.01	10 min	13.00	77 min	5.99	0.20	11.65	0.39
50	8506.49	86 min	89.58	10 min	10.38	96 min	7.38	0.15	14.53	0.29
100	12030.00	122 min	92.42	10 min	7.57	132 min	10.03	0.10	19.98	0.20

*Sampling cost per 49 samples.

3 Chapter 2. Models for predicting the seasonal dynamics of the whitefly *Bemisia tabaci* in soybean crops through neural networks.

3.1 Abstract

Bemisia tabaci (Gennadius 1889) (Hemiptera: Aleyrodidae) is a primary pest in essential crops for food security, such as tomato, potato, brassicas, and soybean. Soybean (*Glycine max*) (L) (Merr) is the leading source of animal protein in the world and the first oilseed grown worldwide. *B. tabaci* remains a relevant pest in soybean crops; on average, one whitefly per sample yields 30 kg ha^{-1} of losses. Determining the seasonal dynamic of *B. tabaci* is helpful to control the pest at the right time, avoiding losses. This investigation aims to model the seasonal dynamics of *Bemisia tabaci* in soybean crops through neural networks. This research tracked whitefly density, climatic elements, and soybean age in 100 soybean fields in the Brazilian cerrado to build seasonal dynamic models from 2017 to 2019. Features were selected according to correlation analysis and biological meaning. ANNs structures were investigated to forecast whitefly through the years and the model presenting the highest *Pearson* correlation and lowest root mean square error were chosen. Feature importance was analyzed to examine these attributes' effect on *B. tabaci*. Then, the model was validated by comparing whitefly observed and fit data densities in the period of study. The ANN selected has five entries (soybean age, average temperature, rainfall, wind speed, and atmosphere pressure) and four neurons in the hidden shell. Average temperature and wind speed are key features in the model presenting the most elevated relative importance index to predict whitefly adult population. Therefore, this study highlighted artificial intelligence's power in modelling a key pest's seasonal dynamic upon a range of attributes seven days in advance.

Keywords: artificial intelligence, ANN, forecast, *Glycine max*.

3.2 Introduction

Insect pest populations are regulated by biotic and abiotic forces in crop systems; These forces would determine the pest attack magnitude throughout time (Chamuene et al., 2020; Prasad, K.V.H., 2022). Understanding these factors is essential for forecasting pest attacks and planning control actions. However, a plenty of variables act on these arthropods making this modeling complex. Despite that, a dataset collected in realistic crop systems associated with artificial intelligence tools would be appropriate to investigate seasonal pest dynamics (Aparecido et al., 2020; Farias et al., 2022).

Artificial intelligence (AI) is any technique that enables a computer to mimic human behavior (Amini, 2019). AI methods such as machine learning algorithms and artificial neural networks (ANN) are adopted to solve regression, classification, clustering, and dimensionality reduction issues in extensive areas of science (Ertel, 2018; Amini, 2019). The computer power, data availability, and AI libraries in open access platforms such as Python (e.g., scikit-learn, keras) and R software expanded their usage to a broad community nowadays (Amini, 2019; Fradkov, 2020). According to the Pubmed database, in 2011, 5250 papers containing the term "artificial intelligence" were published; otherwise, ten years later, 30865 studies were published (almost six times more) (Pubmed, 2022). Besides the medical area, AI tools have been applied in Entomology studies to identify, count, classify, and forecast insect populations, and so forth (Hong et al., 2021; Hoye, et al., 2021; Pubmed, 2022).

ANN is an AI branch able to deal with complex databases containing a range of entries (Amini, 2019). When this technique is applied to a robust dataset through a training and validated process, the network would be able to estimate and preview factors in a study with high accuracy (Benos et al., 2021; Skawsang et al., 2019; Mosaffaei et al., 2021).

A pest that demands attention in agriculture worldwide is the whitefly *Bemisia tabaci* (Gennadius 1889) (Hemiptera: Aleyrodidae) because it may lead to billions of dollars in losses. This pest reduces crop yield by sucking the phloem sap, injecting toxins, transmitting viruses, and promoting soft molding growth (Byrne & Bellows, 1991; Li et al., 2021). Besides, it is challenging to manage due to their fast reproduction, especially in tropical lands, and *B. tabaci* populations developed resistance to extensive insecticide classes globally (165 counties) (Willis, 2017; Horowitz et al., 2020). *Bemisia tabaci* is a polyphagous pest that feeds on over 1000 plant species, being a primary pest in some essential crops for food security, such as tomato, potato, brassicas, and soybean (Jones, 2003; Li et al., 2021).

Soybean (*Glycine max*) (L) (Merr) is a primary source of animal protein in the world and the first oilseed grown worldwide. Brazil, the United States of America, Argentina, China, and India are responsible for 86.49% of global soybean production (USDA, 2022). In 2020, 373 million tons of *G. max* grains were produced globally, and over 176 billion dollars were traded (importation + exportation) across the planet (Faostat, 2020). Insect pest plays a role in soybean operations; these arthropods reduce soybean production by 25% (Gaur & Mogalapu, 2018). *B. tabaci* remains a relevant pest in soybean crops; on average, one whitefly per sample yields 30 kg. ha^{-1} of losses (Padilha et al., 2021).

Among the factors that govern the seasonal dynamic of *Bemisia tabaci* in soybean crops, climatic elements (temperature, humidity, wind, e.g.) and soybean age are potential candidates (Naranjo et al., 2009; Ramos et al., 2019; Pathania et al., 2020). These features are directly attached to *B. tabaci* biology and ecology (Ramos et al., 2018; Sani et al., 2020). Although the relationship among these features is complex, achieving practical knowledge is challenging. AI analyses would potentially modeling the seasonal dynamic of *B. tabaci* in soybean crops through a wide range of variables, yielding valuable information (Farias et al., 2022). Then, this study would elucidate factors that mainly regulate *B. tabaci* in this environment and forecast this pest density for planning management tactics. Consequently, this investigation aims to determine a prime model for predicting the seasonal dynamics of *Bemisia tabaci* in soybean crops through neural networks. In this research, whitefly density, climatic elements, and plant age were tracked in soybean fields to build seasonal dynamic models to forecast whitefly in tropical soybean crops.

3.3 Material and methods

3.3.1 Material and methods

In this research, whitefly density, climatic elements, and soybean age were tracked in *G. max* fields to build seasonal dynamic models to forecast whitefly in tropical soybean crops.

3.3.2 Study site

This study was performed in 100 commercial soybean sites in Formoso do Araguaia (11°55'23.80"S; 49°41'40.70"O, 240 m altitude) and Gurupi (11°48'10.30"S; 49°00'29.30"O, 287 m altitude), state of Tocantins, Brazil from 2017 to 2019. The soybean variety M 8808 IPRO (Intacta) (Agro Bayer Brasil) was sowed at 0.45 m between rows, 0.02 m deep, and 13 seeds per meter, yielding 288,888.89 plants per crop.

3.3.3 Data acquisition

Whitefly adults and nymphs were scouted in 25 plants (regular grid) per plot in the previously described fields. Whitefly adults were evaluated by beating the soybean apex against a plastic tray and counting the number of insects on the tray button. Nymphs were evaluated in the same plants, although they were direct counting in a leaf in the middle section of the plant canopy. Plant age (days after sowing) (DAT) have been tracked on each whitefly density evaluation. Finally climatic element data were obtained across crop development in a meteorological station near the sites.

3.3.4 Dataset

The earliest dataset to forecast whitefly seasonal dynamic in tropical soybean crops comprised a hundred rows and nine columns. This dataset consisted of seven independent variables (features) and two dependent variables (y). The features evaluated in this study were, soybean age (DAT), and climatic element data. These features were selected because they relate to whitefly biology and ecology. The dependent variables were whitefly adults and nymphs. These variables are detailed in the following paragraphs.

In this study, the following climatic element were tracked through the crop seasons: wind speed (m.s^{-1}), rainfall (mm. day^{-1}), relative humidity (%), solar irradiation (J.m^{-2}), average temperature (°C), and atmospheric pressure (KPa). Climatic elements play a role in regulating insect's metabolism, leading direct (raining drops impact and flooding, e.g.) or indirect (favor fungi parasites growth) mortality to pest population (Sharma, & Yogesh 2014; Marabi et al., 2017). Soybean age was examined because it would be correlated with pest population dynamic.

Whitefly adults and nymphs were initially checked as independent variables in the model. These phases represent most of the whitefly population structure. Therefore, they were incorporated into the primary dataset.

3.3.5 Feature selection

Pearson correlation analysis among the independent variables was performed to eliminate correlated features.

3.3.6 Training the models

All the models and analysis were made in R software (R version 3.6.2 2019-12-12). First, data were randomly split in the training (70%) and test (30%) dataset by the function *sample.split* on *caTools* package (Tuszynski 2019). Also, data were centered by subtracting the features by the mean and scaled to facilitate model conversion from function *scale*, package *stats*. Test data were kept separated and assessed just in the validation processes. Models were built by function *mlp*, package *RSNNS* package (Bergmeir and Bemítez 2012). ANN with learning (R prop and Standard Backpropagation) and activation function were tested (Hyperbolic tangent and Logistic). Besides, ten ANN sizes (number of neurons in the hidden layer) and lagged periods to whitefly adults and nymphs were examined to select the best model to study *B. tabaci* seasonal dynamic in soybean crops. Then, 100 interactions were fixed, and initial random weight was generated by the function initFunc = “Randomize_Weights”. Finally, a total of 800 ANN was yield (2 learning functions x 2 activation function x 10 neurons size x 2 independent features x 10 intervals of lagged periods).

3.3.7 Validating the models

Models were validated comparing observed against predicted data by ANN fitted. Pearson’s correlation (rv) and root mean square error (RMSE) were calculated for the validated dataset. The selected and most accurate ANN model would be the one with the highest rv and lowest RMSE values. After, relative importance of the variables was estimated through Oldenn’s algorithm (function *olden*, package *NeuralNetTools*) (Olden et al, 2004; Costa et al 2021). To calculate 95% confidence (CI $_{95}$), the selected ANN were running 250 times, starting with random weights and Olden’s values were recorded and combined to yield the CI$_{95}$ (Beck 2018). Lastly, sensitivity analyses were adopted using partial dependence plots (function *partial*, package *pdp*) to show the relationship between an independent feature with a predictor assuming the average effect to the others variables (Greenwell 2017).

3.4 Results

3.4.1 Feature selection

Some weather features were correlated; rainfall and relative humidity (r= 0.61) as temperature average and solar radiation (r=0.57) presented correlation. Also, the wind speed correlated with relative humidity (r= -0.89). Then, one of these entries was adopted in the model to simplify it (Table 1). Finally, average temperature, rainfall, wind speed, and DAT were selected to compose the models for seasonal dynamic of *B. tabaci* in tropical soybean crops (Table 1, Figure 1).

3.4.2 ANN selection

The ANN selected (ANN setting 2) to model the seasonal dynamic of *Bemisia tabaci* adults in soybean crops was constituted of the following setting: 7 days of lag, four neurons in the hidden shell, Rprop as learning function, and logistic as learning function. This ANN was selected because it presented the highest Pearson correlation coefficient between fitted and observed data (rv= 0.960) and the lowest RMSE (0.091). All the ANN tested to model the seasonal dynamic of *Bemisia tabaci* nymphs showed lower rv and higher RMSE than the previously selected ANN for *B. tabaci* adults (Table 2). Also, whitefly nymphs are more laborious to scout than adults. Then, ANN setting 2 was assigned for predicting the seasonal dynamic of the whitefly *Bemisia tabaci* in soybean crops. Therefore, the most suitable ANN structure has five entries (DAT, average temperature, rainfall, wind speed, and atmosphere pressure), four neurons in the hidden shell, and a whitefly adult's density as output (Figure 1, table 2).

3.4.3 Whitefly adult's seasonal dynamic

During the study, the whitefly population was high as the average air temperature increased even when rain occurred. Although in year three, precipitation down-regulated whitefly density in high temperature condition. Atmosphere pressure did not change abruptly across the years and did not shape the whitefly population much. However, wind speed varied across the time and seems down-regulated whitefly. In year two, for example, the peak of whitefly was higher than in year one, and the wind speed was lower when *B. tabaci* achieved this prominent density peak in February of year two (Figure 2).

3.4.4 Feature relative importance

Average temperature, and wind speed presented the most elevated relative importance index to forecast whitefly adult population, followed by DAT, rainfall, and atmospheric pressure. Average temperature, wind speed and DAT had affected *B. tabaci* density positively. Whereas rainfall and atmospheric pressure affected the whitefly population negatively (Figure 3). The number of *B. tabaci* adults alters when an attribute varies while others is fixed on average, reinforcing that each feature of the selected ANN has a relative dependence on estimating whitefly density (Figure 4).

3.4.5 ANN Validation

The selected ANN was able to preview the whitefly adult population with seven days of lag. This fact is demonstrated by the match between the observed whitefly population across the years of study (June 2017 to April 2019) and the fitted population in the same period by the selected ANN (Figure 5). Therefore, the convergence of whitefly field population and fitted by the model, validated the capability of the ANN in forecast *B. tabaci* population seven days in advance.

3.5 Discussion

The artificial neural network selected in this study could model the seasonal dynamic of the whitefly adult population in soybean crops through climatic elements and plant age.

Plenty of features would determine the seasonal dynamic of whitefly in soybean fields. Then, selecting key factors to model this dynamic is mandatory to avoid multicollinearity and complex models. Correlation analyses have been a classical tool to remove unnecessary attributes. Also, biological meaning and the researcher's expertise are essential to building these regression models. Following this guide, we have selected five entries to the models. Four of them were whether elements and the other was soybean age. These features shape the biology and ecology of whitefly in soybean fields.

Artificial neural networks have been adopted to model complex regression studies in science. The structure of the selected ANN has five entries, four neurons in a hidden layer, and one exit layer. The learning algorithm Rprop updates ANN weights using gradient signs and distinct step sizes per synaptic weight, increasing training performance. Logistic function makes it possible to shape the data in a non-linear manner, improving the generalization power of the ANN. A simple ANN structure was achieved due to the effective feature selection and training process adopted during the investigation. This modeling better forecasts whitefly adults with seven days of lag rather than nymphs. Scouting whitefly adults is less laborious than nymphs, enabling its application in research and integrated pest management programs. The seven days of lag would be modeling this insect effectively, allowing action before a pest outbreak.

The seasonal dynamic of whitefly in soybean fields is regulated by a range of biotic and abiotic factors. However, this study showed that some primary entries regulate this seasonal dynamic, presenting biological meaning.

Climate elements are critical to insect biology and ecology. The minimum temperature threshold for whitefly completes their entire life cycle (egg to adult) is 8.7 to 15 ºC (Bradshaw et al., 2019). Also, the optimum temperature for whitefly development and survival is 20 to 30 ºC. Besides, the minimum temperature for *B. tabaci* (Biotype Q) oviposition is 14 ºC, while for flight is 19.2 to 22.0 ℃ (Bradshaw et al., 2019; Enkegaard, 1993). This temperature thresholds evidence the importance of average air temperature as a feature in modeling the seasonal dynamic of *Bemisia tabaci* (Enkegaard, 1993; Bradshaw et al., 2019). This study converges with the effect of temperature on whiteflies, increasing their population as the temperature rises (Marabi, Bhowmick, Pachori, & Sharma, 2017). According with the model, wind speed positively affected whitefly adults, possibly favoring migration, and mating processes. Although it was not clear observed in the whitefly field fluctuation data maybe because there was not enough variation in the wind speed. However, rainfall acts negatively against the whitefly, reducing its density. Raindrops' impact causes direct insect mortality (Marabi, Bhowmick, Pachori, & Sharma, 2017; Sharma, & Yogesh 2014).

Average temperature, wind speed, rainfall, and soybean age were critical features to modeling whitefly seasonal dynamics as these elements exhibited the most significant feature importance. Climatic factors are known to regulate insects' metabolism and fitness. Temperature regulates insect development rate affecting female adults' survival time, fecundity, and immature development time, and affects the number of insects yields per crop season (Aregbesola et al., 2020). *B. tabaci* Cassava populations have shown a reduction in adult survival time of 20 days to 8 days in 24 °C to 36 °C temperatures, respectively. Also, the temperature of 16 °C promotes an immature nymph stage duration of 59 days, conversely at 28 °C immature time was 16 days (Aregbesola et al., 2020). Wind speed affect insect migration and colonization on crops; wind speed would facilitate insect dispersion; or even interrupts it, depending on wind direction (Lima et al., 2018; Walerius et al., 2023). Rainfall generally harms sucking pests directly; rainfall drops mechanically damage *B. tabaci* eggs and immatures due to the impact of the drop on the insect body and indirectly by increasing the humidity and favoring entomologic fungi action on this Hemiptera (Naranjo and Ellsworth, 2005; Felicio et al., 2019; Sani et al., 2020). Finally, whitefly population tends to progress across plant development by yielding novel cycles in the crop and being more protected against pesticides as plant rows canopy close.

The selected ANN modelled the whitefly seasonal dynamic with high precision. Observed and predicted *Bemisia tabaci* density adults matched across the years. This ANN was able to forecast whitefly when their population is increasing or decreasing in distinct periods. It highlights ANN's power in predicting insect pests' seasonal dynamic throughout a complex feature structure (Oliveira Aparecido et al., 2020). ANN are among the main tools to solve complex regression problems; this technique was essential to understanding the features and modelling the seasonal dynamic of whitefly with towering precision.

3.6 Conclusion

The artificial neural network selected in this study is reliable for modelling the seasonal dynamics of the whitefly *Bemisia tabaci* in soybean crops. The ANN structure has five entries (DAT, average temperature, rainfall, wind speed, and atmosphere pressure) and four neurons in the hidden shell. This model forecasts whitefly adults with high accuracy and seven days in advance using Rprop and logistic function as learning algorithm and activation function, respectively. Average temperature and wind speed play a role in the model presenting the most elevated relative importance index to forecast the whitefly adult population. This study highlighted artificial intelligence's power in modelling a key pest's seasonal dynamic upon a realistic range of features in seven days of advance. Finally, this technique would be expanded to other pest systems and devices to add value to monitoring and management tools.

3.7 Acknowledgments

Financial support was provided by the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq); Coordination of Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Finance Code 001); and Minas Gerais State Foundation for Research Aid (Fundação de Amparo a Pesquisa do Estado de Minas Gerais– FAPEMIG).

3.8 References

Amini, A. (2019). Introduction to deep learning. MIT, 6, S191.

Chamuene, A., Araújo, T.A.D., Lopes, M.C., Ramos Pereira, R., Berger, P.G., & Picanço, M.C. (2020). Investigating the natural mortality of *Aphis gossypii* (Hemiptera: Aphididae) on cotton crops in tropical regions using ecological life tables. *Environmental entomology*, *49*(1), 66-72.

Aparecido, L.E.O., Rolim, G.S., Moraes, J.R.C., Costa, C.T.S., & Souza, P.S. (2020). Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, *64*(4), 671-688.

Aregbesola, O. Z., Legg, J. P., Lund, O. S., Sigsgaard, L., Sporleder, M., Carhuapoma, P., & Rapisarda, C. (2020). Life history and temperature-dependence of cassava-colonising populations of *Bemisia tabaci*. *Journal of Pest Science*, *93*, 1225-1241.

Beck, M.W (2018). NeuralNetTools: visualization and analysis tools for neural networks. *Journal of Statistical Software*, *85*:1–20.

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Bergmeir, C.N., & Benítez Sánchez, J.M. (2012). Neural networks in R using the Stuttgart neural network simulator: RSNNS. *Journal of Statistical Software*, *46*,1–26.

Byrne, D.N., & Bellows, T.S. (1991). Whitefly biology. *Annual review of entomology*, *36*, 431–457.

Costa, N.L., Lima, M.D., & Barbosa, R. (2021). Evaluation of feature selection methods based on artificial neural network weights. *Expert Systems with Applications*,*168*, 114312.

Enkegaard, A. (1993). The poinsettia strain of the cotton whitefly, *Bemisia tabaci* (Homoptera: Aleyrodidae), biological and demographic parameters on poinsettia (Euphorbia pulcherrima) in relation to temperature. *Bulletin of Entomological Research*, *83*(4), 535–546.

Ertel, W. (2018). Introduction to artificial intelligence. Springer.

Farias, E.S., Farias, A.A., Santos, R.C., Santos, A.A., & Picanço, M.C. (2022). Forecasting the seasonal dynamics of *Trichoplusia ni* (Lep.: Noctuidae) on three Brassica crops through neural networks. *International Journal of Biometeorology*, *66*(5), 875–882.

Felicio, T.N.P., Costa, T. L., Sarmento, R. A., Ramos, R. S., Pereira, P. S., Silva, R. S., & Picanço, M.C. (2019). Surrounding vegetation, climatic elements, and predators affect the spatial dynamics of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in commercial melon fields. *Journal of Economic Entomology*, *112*(6), 2774–2781.

Fradkov, A.L. (2020). Early history of machine learning. IFAC-PapersOnLine, *53*(2), 1385-1390.

Greenwell, B.M (2017) pdp: an R package for constructing partial dependence plots. R J 9:421–436

Hong, S.J., Nam, I., Kim, S.Y., Kim, E., Lee, C.H., Ahn, S., ... & Kim, G. (2021). Automatic pest counting from pheromone trap images using deep learning object detectors for *Matsucoccus thunbergianae* monitoring. *Insects*, *12*(4), 342.

Horowitz, A.R., Ghanim, M., Roditakis, E., Nauen, R., & Ishaaya, I. (2020). Insecticide resistance and its management in *Bemisia tabaci* species. Journal of Pest Science, *93*(3), 893–910.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(3), 346–363.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., & Hothorn, M.T. (2016). Package 'multcomp'.

Høye, T.T., Ärje, J., Bjerge, K., Hansen, O.L., Iosifidis, A., Leese, F., ... & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, *118*(2), e2002545117.

Jones, D.R. (2003). Plant viruses transmitted by whiteflies. *European journal of plant pathology*, *109*(3), 195–219.

Li, Y., Mbata, G.N., Punnuri, S., Simmons, A.M., & Shapiro-Ilan, D.I. (2021). *Bemisia tabaci* on vegetables in the Southern United States: incidence, impact, and management. *Insects*, *12*(3), 198.

Lima, C.H., Sarmento, R.A., Galdino, T.V., Pereira, P.S., Silva, J., Souza, D.J., ... & Picanço, M. C. (2018). Spatiotemporal dynamics of whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in commercial watermelon crops. *Journal of economic entomology*,*111*(4), 1895–1903.

Marabi, R. S., Das, S. B., Bhowmick, A. K., Pachori, R., & Sharma, H. L. (2017). Seasonal population dynamics of whitefly (*Bemisia tabaci* Gennadius) in soybean. *Journal of Entomology and Zoology Studies*, *5*(2), 169–173.

Mosaffaei, Z., & Jahani, A. (2021). Modeling of ash (*Fraxinus excelsior*) bark thickness in urban forests using artificial neural network (ANN) and regression models. *Modeling Earth Systems and Environment*, *7*(3), 1443–1452.

Naranjo, S. E., & Ellsworth, P. C. (2005). Mortality dynamics and population regulation in Bemisia tabaci. *Entomologia experimentalis et Applicata*, *116*(2), 93–108.

Naranjo, S.E., Castle, S.J., Barro, P.J., Liu, S.S. (2009). Population Dynamics, Demography, Dispersal and Spread of *Bemisia tabaci*. In: Stansly, P., Naranjo, S. (eds) Bemisia: Bionomics and Management of a Global Pest. Springer, Dordrecht.

Olden, J.D., Joy, M.K., & Death, R.G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological modelling*, *178*(3-4), 389–397.

Oliveira Aparecido, L.E., Souza Rolim, G., Silva Cabral De Moraes, J.R., Costa, C.T. S., & Souza, P.S. (2020). Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, *64*, 671–688.

Pathania, M., Verma, A., Singh, M., Arora, P.K., & Kaur, N. (2020). Influence of abiotic factors on the infestation dynamics of whitefly, *Bemisia tabaci* (Gennadius 1889) in cotton and its management strategies in North-Western India. *International journal of tropical insect science*,*40*, 969–981.

Prasad, K.V.H. (2022). Factors Affecting Growth and Metabolism of Insects. In: *Insect Ecology*: *Concepts to Management*. Springer, Singapore.

Ramos, R. S., Kumar, L., Shabani, F., & Picanço, M. C. (2018). Mapping global risk levels of *Bemisia tabaci* in areas of suitability for open field tomato cultivation under current and future climates. PloS one, 13(6), e0198925.

Ramos, R.S., Kumar, L., Shabani, F., Silva, R.S., Araújo, T.A., & Picanço, M.C. (2019). Climate model for seasonal variation in *Bemisia tabaci* using CLIMEX in tomato crops. *International journal of biometeorology*, *63*, 281–291.

Sani, I., Ismail, S. I., Abdullah, S., Jalinas, J., Jamian, S., & Saad, N. (2020). A review of the biology and control of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae), with special reference to biological control using entomopathogenic fungi. *Insects*, *11*(9), 619.

Sharma, S. S., & Yogesh, K. (2014). Influence of abiotic weather parameters on population dynamics of whitefly, *Bemisia tabaci* (Genn) on cotton. *Journal of Cotton Research and Development*, *28*(2), 286-288.

Skawsang, S., Nagai, M., K. Tripathi, N., & Soni, P. (2019). Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: a case study for the Central Plain of Thailand. *Applied Sciences*, *9*(22), 4846.

Tuszynski J (2019) caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. https://cran.r-project.org/package=caTools. Accessed 30 may 2023.

Walerius, A.H., Pallini, A., Venzon, M., Santana Júnior, P.A., Costa, T.L., Paes, J.D. S., ... & Picanço, M.C. (2023). Use of Geostatistics as a Tool to Study Spatial-Temporal Dynamics of *Leucoptera coffeella* in Coffee Crops. *Agriculture*, *13*(2), 438.

Willis, K. (2017). State of the world's plants 2017. Royal Botanics Gardens Kew.

3.9 Figures and tables

Table 1. Pearson's correlation coefficients of weather variables among themselves.

Atmospheric pressure	0.16	0.38	-0.19	-0.11	-0.21
Rainfall		0.61	-0.20	-0.22	-0.45
Relative humidity			-0.38	-0.33	-0.89
Solar irradiation				0.57	0.18
Temperature (average)					0.16
	Rainfall	Relative humidity	Solar irradiation	Temperature (average)	Wind speed

Table 2. Topology (number of neurons, activation function, and learning algorithm) and performance (rv = Pearson's correlation of fitted and observed data and RMSE = mean square error in the validation set) of artificial neural networks (ANN) designed to predict the seasonality of the whitefly *Bemisia tabaci* in soybean crops.

y	§ Lag	Neurons	Activation function	Learning function	rv	RMSEv
Adults	0	4	Logistic	Rprop	0.88079	0.13438
	7*	4	Logistic	Rprop	0.96043	0.09078
	14	7	Sigmoid	Rprop	0.90509	0.09274
	21	11	Hyperbolic tangent	SCG	0.77281	0.19262
	28	5	Hyperbolic tangent	SCG	0.84001	0.11545
	35	9	Sigmoid	Rprop	0.85203	0.12004
	42	7	Sigmoid	Rprop	0.80472	0.1385
	49	5	Hyperbolic tangent	Rprop	0.79647	0.15156
	56	8	Sigmoid	Rprop	0.88669	0.10422
	63	9	Logistic	Rprop	0.87303	0.11101
Nymphs	0	4	Logistic	Rprop	0.33979	0.21698
	14	5	Sigmoid	Rprop	0.68091	0.2199
	21	7	Logistic	Rprop	0.50769	0.19838
	28	5	Hyperbolic tangent	Rprop	0.68755	0.1731
	35	4	Logistic	Rprop	0.64404	0.19417
	42	3	Hyperbolic tangent	SCG	0.66289	0.20719
	49	11	Sigmoid	Rprop	0.49726	0.21806
	56	6	Logistic	Rprop	0.54212	0.1968
	63	11	Logistic	Rprop	0.60576	0.1847

§ Number of days before the evaluation of *B.tabaci* intensity attack in which the mean data for the meteorological variables were calculated.

*Model selected due higher Pearson correlation and lower mean square error.

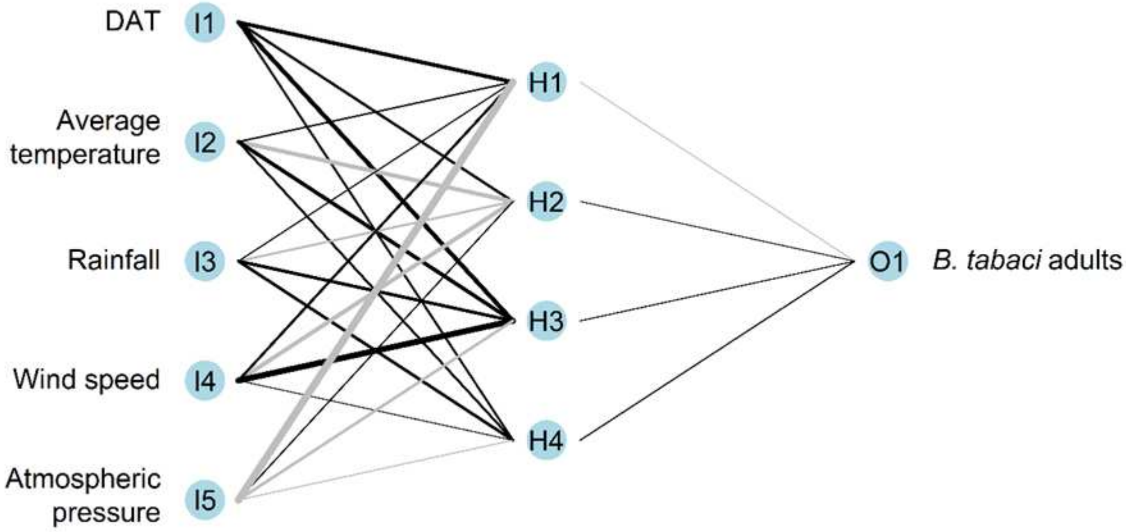

Figure 1. Diagram of the winning artificial neural network to predict the attack intensity of *Bemisia tabaci* adults on soybean crops. This artificial neural network was defined by five predictors (I1 to I5) in the input layer, four neurons (H1 to H4) in the hidden layer and one neuron (O1) in the output layer. I1 = age of soybean plants (days), I2 to I5 = average data of meteorological variables from the last seven days: I2 = air temperature (oC), I3 = rainfall (mm.day^{-1}), I4 = average wind speed (m.s^{-1}) and I5 = atmospheric pressure (KPa). The black and gray lines represent positive and negative weights between layers, respectively. The thickness of the line corresponds to the relative magnitude of each weight.

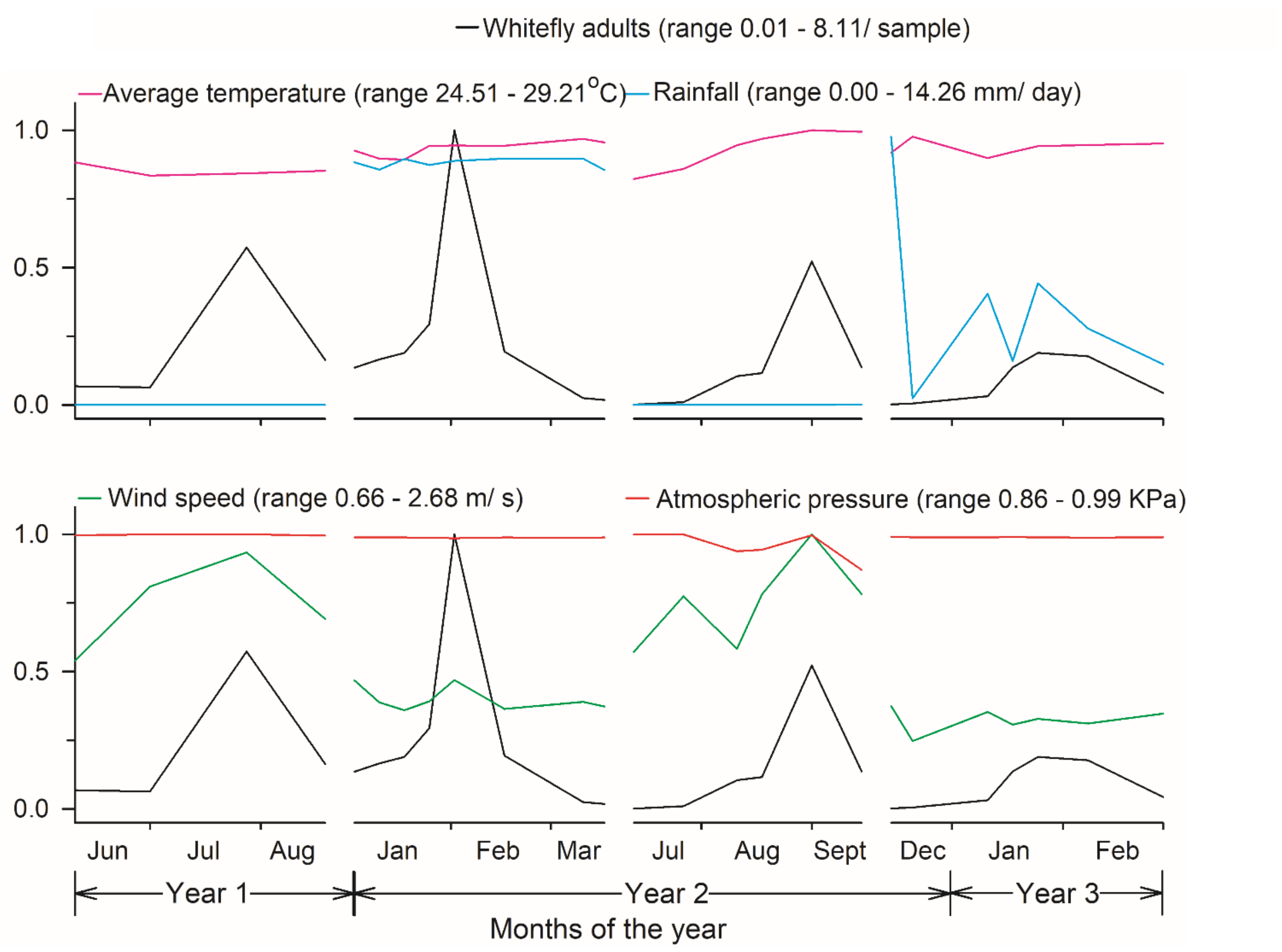

Figure 2. Seasonal variation of whitefly *Bemisia tabaci* adults desities and meteorological variables (with a 7-day delay in the evaluation of pest densities) mean air temperature, rainfall, wind speed, and atmospheric pressure during the period of this work.

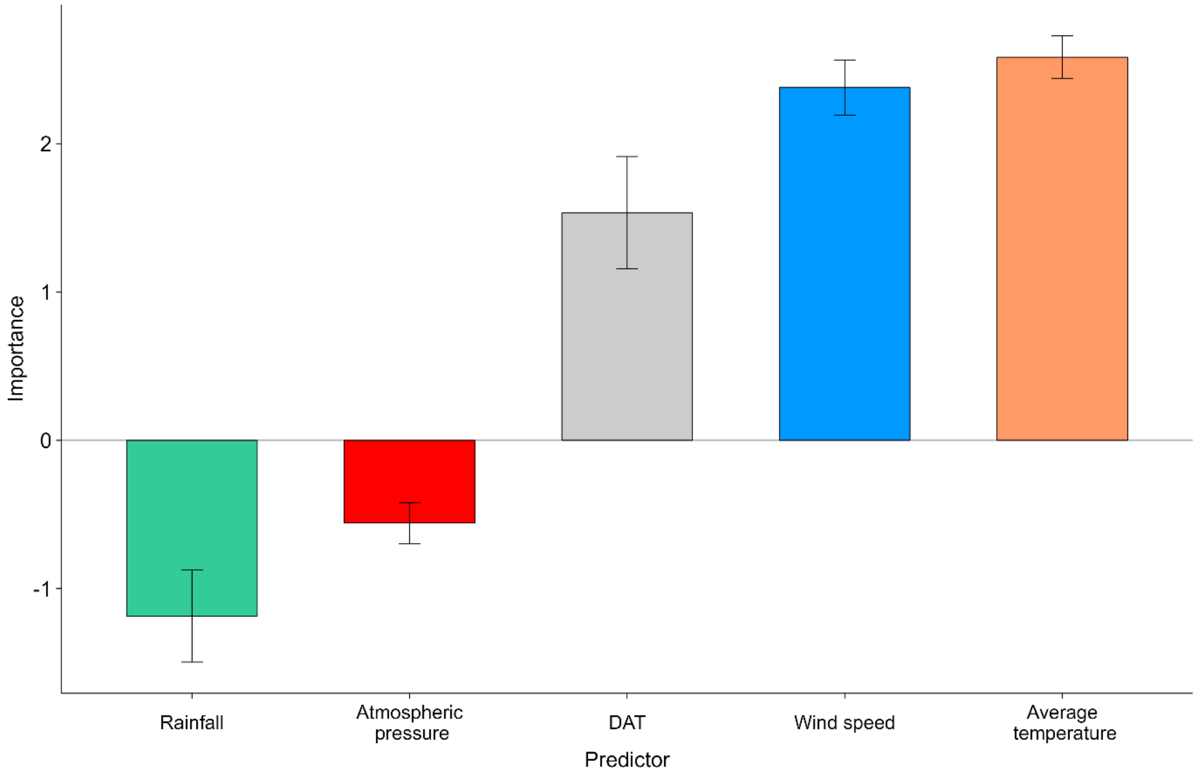

Figure 3. Relative importance (mean and 95% confidence interval) of the predictors for the response variable estimated from the weights of the winning artificial neural network (ANN) by Olden's method. 250 ANNs were created with different initial weights, and the importance of the predictors was calculated and stored to allow the calculation of the mean and confidence intervals.

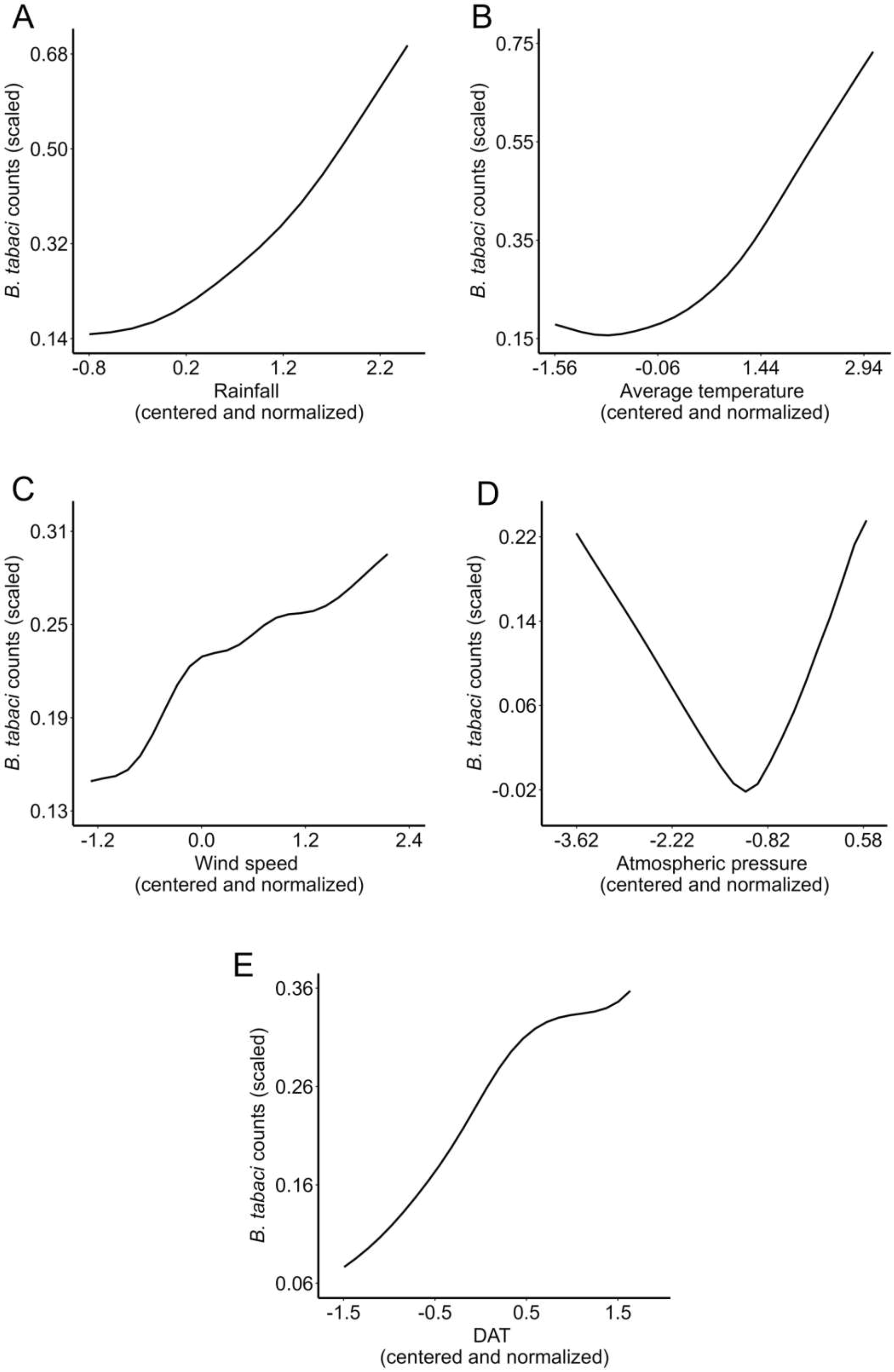

Figure 4. Partial dependence of the winning network showing the average output variation (density of whitefly *Bemisia tabaci* adults) in relation to the levels of the input variables: (A) rainfall schedule, (B) average air temperature, (C) average wind speed, (D) atmospheric pressure and (E) soybean plants age.

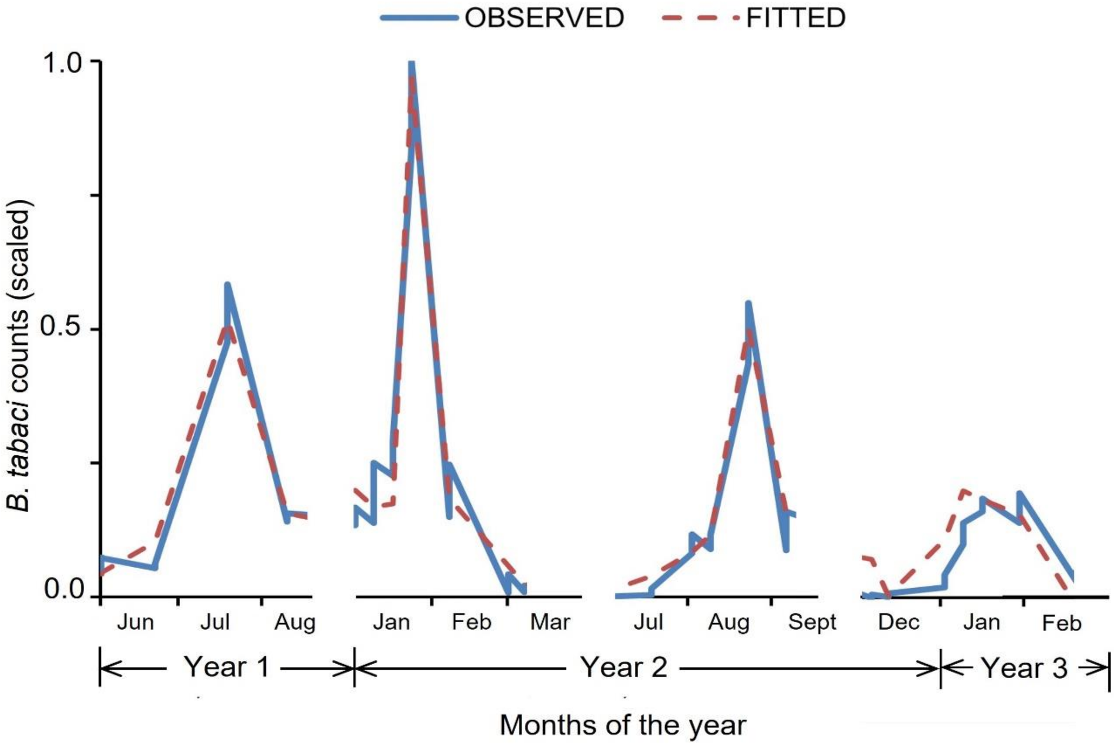

Figure 5. Forecated and observed densities of whitefly *Bemisia tabaci* adults in Brazilian cerrado fields during the period of this study.

4 General conclusion

Whitefly is a primary issue in soybean crops, especially in dry and hot areas across tropical lands. Its attack can reduce yield in 1000 kg per ha. This study provides technical guidance about whitefly *Bemisia tabaci* scouting and seasonal dynamic in soybean tropical crops. In the first chapter, a feasible sampling plan for scouting *B. tabaci* was developed and in the second chapter, a seasonal dynamic model using artificial neural network was developed. The conventional sampling is precise, fast, and representative. It consists in evaluate forty-nine soybean plants, assessing whitefly adults by beating a plastic tray against the apical part of plant canopy. Seasonal dynamics of *B. tabaci* in soybean was modeling by an ANN consisting in five entries (DAT, average temperature, rainfall, wind speed, and atmosphere pressure) and four neurons in the hidden shell. Upon this model, the whitefly adult population is forecasted seven days in advance with high accuracy. Average temperature, and wind speed are critical components in the model presenting the most elevated relative importance index to forecast the whitefly adult population. Finally, this investigation provides relevant accomplishment about whitefly manage in soybean neotropical area. This sampling plan and seasonal dynamic study facilitate growers identify early pest infestation in the field and plan control tactics before this pest cause economical damage. Future works must develop sequential sampling plans and implement this knowledge in Integrate Pest Manage digital tools.