Economic-threshold-based classification of soybean aphid, *Aphis glycines*, infestations in commercial soybean fields using Sentinel-2 satellite data

Arthur V. Ribeiro[a,*], Lorena N. Lacerda[b], Marcella A. Windmuller-Campione[c], Theresa M. Cira[a,1], Zachary P. D. Marston[a,2], Tavvs M. Alves[a,3], Erin W. Hodgson[d], Ian V. MacRae[e], David J. Mulla[f], Robert L. Koch[a]

[a] Department of Entomology, University of Minnesota, 1980 Folwell Avenue, Saint Paul, MN, 55108, USA

[b] Department of Crop and Soil Sciences, University of Georgia, 3111 Miller Plant Science Building, 120 Carlton Street, Athens, GA, 30602, USA

[c] Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue North, Saint Paul, MN, 55105, USA

[d] Department of Plant Pathology, Entomology and Microbiology, Iowa State University, 2213 Pammel Drive, Ames, IA, 50011, USA

[e] Department of Entomology, University of Minnesota, Northwest Research and Outreach Center, 2900 University Avenue, Crookston, MN, 56716, USA

[f] Department of Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, Saint Paul, MN, 55108, USA

[1] Present address: Pesticide and Fertilizer Management Division, Department of Agriculture, 625 Robert Street North, Saint Paul, MN, 55155, United States

[2] Present address: Syngenta, 9 Davis Drive, Research Triangle Park, NC, 27709, USA

[3] Present address: Instituto Federal de Educação, Ciência e Tecnologia Goiano, Campus Rio Verde, Rodovia Sul Goiana km 01 Zona Rural, Rio Verde, GO, 75901-970, Brazil

* Corresponding author at: Department of Entomology, University of Minnesota, 1980 Folwell Avenue, Saint Paul, MN, 55108, USA

E-mail addresses: vieir054@umn.edu (A.V. Ribeiro), llacerda@uga.edu (L.N. Lacerda), mwind@umn.edu (M.A. Windmuller-Campione), theresa.cira@state.mn.us (T.M. Cira), zach.p.marston@gmail.com (Z.P.D. Marston), tavvs.alves@ifgoiano.edu.br (T.M. Alves), ewh@iastate.edu (E.W. Hodgson), imacrae@umn.edu (I.V. MacRae), mulla003@umn.edu (D.J. Mulla), koch0125@umn.edu (R.L. Koch)

Abstract

The soybean aphid (SBA), *Aphis glycines* Matsumura (Hemiptera: Aphididae), is a significant insect pest of soybean, *Glycine max* (L.) Merrill (Fabales: Fabaceae), and field treatment decisions for this pest are based on average field populations. Previous studies indicated that ground- and drone-based red-edge and near-infrared remote sensing can be used to detect plant stress caused by SBA infestations in soybean. However, it remains to be determined if remote sensing for SBA can be expanded to field or landscape scale using satellite-based platforms. Thus, this research was conducted in three steps to determine the potential of using Sentinel-2 satellite data for the classification of SBA infestations in soybean fields using simulated and actual Sentinel-2 satellite spectral reflectance. In the first step, as a proof of concept, hyperspectral data from cage studies were used to simulate Sentinel-2 bands and vegetation indices (VIs), conducted in nine trials at multiple locations between 2013 and 2021. The effects of SBA from caged plants on simulated data were evaluated with random intercept linear mixed models. The satellite simulation indicated a significant effect of SBA on the spectral reflectance of caged soybean plants ($p < 0.05$) for four satellite bands (5, 6, 7, and 8A) and five VIs (NDVI, GNDVI, SAVI, OSAVI, and NDRE). In the second step, actual Sentinel-2 spectral reflectance and corresponding aphid counts of commercial soybean fields, collected from 2017 to 2019, were obtained. The relationship between SBA counts and Sentinel-2 spectral reflectance from commercial soybean fields were evaluated with general linear models. A significant effect of SBA was observed for three satellite bands (6, 7, and 8A) and three VIs (NDVI, SAVI, and OSAVI). In the third step, linear support vector machine (LSVM) models for the classification of SBA infestations as above or below a previously determined economic threshold of 250 aphids per plant were developed using simulated Sentinel-2 bands and VIs from the caged plots, and were tested on actual Sentinel-2 data from commercial soybean fields. The best LSVM model for the classification of aphids in soybean reached 91% accuracy, 85.7% sensitivity, and 93.3% specificity. Thus, simulations with caged plots can be used as an indication of the potential of using satellite data for the detection of plant stresses on a larger scale. Furthermore, this study advances decision-making for SBA, and the developed LSVM model can be used to update regional and local monitoring for the management of SBA.

Keywords: machine learning, linear support vector machine, simulation, soybean aphid

1 Introduction

Pests are a limiting factor for crop production, including soybean, *Glycine max* (L.) Merrill (Fabales: Fabaceae) (Bueno et al., 2021). The soybean aphid (SBA), *Aphis glycines* Matsumura (Hemiptera: Aphididae), is a significant soybean pest, especially in the upper Midwest of the United States (Hesler and Beckendorf, 2021). Aphids are phloem-sucking insects that cause local injury to leaf tissue and systemic disruption of plant physiology (Macedo et al., 2003). Such effects can lead to yield losses due to a decrease in the number of pods, seeds, seed size, and seed quality when SBAs are in high numbers (Ragsdale et al., 2007, 2011).

Traditional management of SBA is performed at a whole-field level and is based on scouting and estimation of SBA density in soybean fields (Hodgson et al., 2004; Ragsdale et al., 2011), so aphids can be treated at an economic threshold (i.e., 250 aphids per plant) to avoid infestations from reaching an economic injury level (i.e., 674 aphids per plant) (Ragsdale et al., 2007, 2011; Koch et al., 2016). Scouting soybean fields for SBA is a time-consuming effort, and the development of new technologies to facilitate field scouting and pest monitoring could increase the adoption of more sustainable management recommendations by farmers (Ragsdale et al., 2011; Bueno et al., 2021).

The development of remote sensing associated with computer processing and information technologies contributes to the advancement of agriculture (Mulla, 2013; Cavaco et al., 2022), particularly for the detection, mapping, monitoring, and management of abiotic and biotic plant stresses, including diseases and insects (Abd El-Ghany et al., 2020; Cavaco et al., 2022; Rhodes et al., 2022). Remote sensing for plant stresses involves the use of contactless sensors to detect the electromagnetic radiation reflected or emitted from plant tissues and relate measures of that radiation to changes in plant physicochemical properties (Mulla, 2013; Abd El-Ghany et al., 2020; Cavaco et al., 2022). Numerous studies have documented the effects of insects on the spectral reflectance of crops and forests using ground-, drone-, and satellite-based sensors (Luo et al., 2013; Santos et al., 2017; Vanegas et al., 2018; Iost Filho et al., 2022; Ma et al., 2023).

Satellites offer greater land coverage than other remote sensing technologies, which might increase the efficiency of field scouting (Rhodes et al., 2022). The use of satellite imagery in agriculture has increased over the last decade with the deployment of equipment with higher spatial and temporal resolution (Mulla, 2013; Rhodes et al., 2022). For example, the Sentinel-2 satellite system is comprised of two nearly identical satellites (Sentinel-2A and B) that offer free-of-charge multispectral imagery from 13 bands (including visible and near-infrared regions of the electromagnetic spectrum) with spatial resolution varying between 10 – 60 m (Table 1), and a revisit frequency of 3–5 days (Drusch et al., 2012). The spectral bands in the visible and near-infrared regions make the Sentinel-2 system especially useful for the characterization of vegetation properties (Drusch et al., 2012; Frampton et al., 2013), including changes caused by insect pests (Hawryło et al., 2018; Abdullah et al., 2019; Prabhakar et al., 2022; Ramos et al., 2022).

Table 1. Characteristics of the multispectral bands of the Sentinel-2 satellites A and B

		Sentinel-2A†		Sentinel-2B†	
Band	Resolution (m)	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)
1	60	442.7	20	442.2	20
2	10	492.7	65	492.3	65
3	10	559.8	35	558.9	35
4	10	664.6	30	664.9	30
5	20	704.1	14	703.8	15
6	20	740.5	14	739.1	13
7	20	782.8	19	779.7	19
8	10	832.8	105	832.9	105
8A	20	864.7	20	864.0	21
9	60	945.1	19	943.2	20
10	60	1373.5	30	1376.9	29
11	20	1613.7	90	1610.4	93
12	20	2202.4	174	2185.7	184

Central wavelength calculated as the barycenter of the spectral response function (ESA, 2015), and bandwidths at full width half maximum as of 21 June 2022 (ESA, 2022)

Development of remote sensing with satellites for plant-pest systems is often facilitated by the simulation of spectral reflectance from ground-based (i.e., proximal) hyperspectral data (D'Odorico et al., 2013; Martins et al., 2017; Abdullah et al., 2019; Osco et al., 2019; Ramos et al., 2022). In particular, the simulation of satellite spectral reflectance and satellite-based vegetation indices (VIs) can be an important step to test the feasibility of using satellite sensors for crop pests of economic importance occurring over extensive areas (Martins et al., 2017; Osco et al., 2019).

Previous studies indicated that proximal and drone-based remote sensing with red-edge and near-infrared regions of the electromagnetic spectrum can be used for the detection of plant stress caused by SBA (Alves et al., 2015, 2019; Marston et al., 2020). More recently, a linear support vector machine (LSVM) model was developed for the classification of SBA on caged soybean plants using proximal remote sensing (Marston et al., 2022). However, it remains to be determined if remote sensing for aphids can be expanded to field- and landscape-scale detection and classification of infestations using satellite-based platforms. Thus, this research was conducted in three steps to determine the potential of using simulated and actual Sentinel-2 imagery for the detection and classification of plant stress caused by SBA infestations in soybean fields. In the first step, as a proof of concept, hyperspectral data from cage studies were used to simulate Sentinel-2 bands and VIs. In the second step, actual Sentinel-2 measurements and corresponding aphid counts of commercial soybean fields were obtained, and the relationship between these two factors was assessed. In the final step, LSVM models for the classification of SBA infestations were developed using simulated Sentinel-2 bands and VIs from the caged plots and were tested on actual Sentinel-2 data from commercial soybean fields.

2 Materials and methods

2.1 Simulation of satellite measurements using caged plots

The ability to use satellite data for the detection of plant stress caused by SBA in soybean fields was first evaluated using simulated Sentinel-2 spectral reflectance and VIs. Simulations were done as described below using ground-based hyperspectral data from cage studies conducted in Minnesota and Iowa, United States.

2.1.1 Caged plots

Field experiments with caged soybean plots were conducted in 2013, 2014, 2017, 2018 and 2021 at the University of Minnesota (UMN) Research and Outreach Center in Rosemount, MN (44.715883° N, 93.097913° W), in 2017 and 2018 at the Iowa State University Northern Research Farm in Kanawha, IA (42.930928° N, 93.792338° W), and in 2019 and 2020 at the UMN Agricultural Experiment Station, Saint Paul, MN (44.9898369° N, 93.1802096° W). The field experiments were conducted similarly in all site-years with the objective of assessing the effect of SBA on soybean spectral reflectance.

Detailed information on planting, infestations, and sampling of trials conducted in 2013 and 2014, and 2017 and 2018 are described in Alves et al. (2015) and Marston et al. (2020), respectively. In short, plots of soybean with an area between 1 and 3.75 m^2 were caged with polyvinyl chloride (PVC) frames covered with white no-see-um mesh (Quest Outfitters, Sarasota, FL, USA) in soybean fields with a seeding rate between 345,000 and 495,000 seeds per ha, and row spacing between 0.17 and 0.76 m. In each year, a total of 11 to 32 cages were established, and populations of SBA were manipulated in each cage with artificial SBA infestations or insecticides to obtain a gradient of infestation. Cages in all locations were artificially infested with 0 to 400 mixed-age (i.e., nymphs + adults) SBA, obtained from a laboratory colony (UMN Saint Paul campus), by manually placing the aphids evenly across the upper canopy of multiple soybean plants. Aphids were transported to the field in a cooler (ice packs at the bottom covered with a cardboard layer to avoid direct contact of the aphids with the ice). Aphid counts were obtained weekly at each site-date with non-destructive sampling by randomly selecting and visually inspecting 5 to 10 plants per cage, and counts were converted to cumulative aphid days (CAD), which is an indication of cumulative plant stress caused by aphids over time (Hanafi et al., 1989; Marston et al., 2020).

Planting, infestations, and sampling of the experiments in 2019, 2020 and 2021 were similar to the previous years (Alves et al., 2015; Marston et al., 2020). Soybean plots had an area of 2.25 m^2 and were caged in soybean fields with a seeding rate of 370,000 seeds per ha and row spacing of 0.76 m. Fields were planted on 16 May (variety Stine ‘13EA12’), 15 May (variety Stine ‘19EA32’), and 15 June (variety Golden Harvest ‘1012E3’) of 2019, 2020 and 2021, respectively. A total of 16 cages arranged in eight blocks were established in the fields in 2019 and 2020, and 12 cages arranged in six blocks in 2021. In each cage, soybean plants were artificially infested with SBA, and weekly aphid counts were obtained from five randomly selected plants and converted to CAD, similarly to the description above. Insecticides were not used to manipulate aphid populations in these three years.

2.1.2 Hyperspectral measurements of caged plots and processing

Hyperspectral measurements (not images) of soybean plants were recorded directly nadir from each cage within 2 h of solar noon with clear sky conditions, or with $<$ 20% cloud cover and a clear view between the sun and the field, to reduce the influence of solar angle and atmospheric effects. Five hyperspectral measurements were taken from each cage after canopy closure using a hyperspectral spectroradiometer with wavelength detection range of 350–2500 ± 3 nm (FieldSpec4 Hi-Res spectroradiometer, ASD Inc., Boulder, CO, USA) in 2013 and 2014, and four to eight measurements per cage with a hyperspectral spectroradiometer with wavelength detection range of 325–

1075 ± 1 nm (FieldSpec® HandHeld 2™ VNIR spectroradiometer, ASD Inc., Boulder, CO, USA) in subsequent years. More details on hyperspectral measurements in 2013 and 2014, and 2017 and 2018 can be found in Alves et al. (2015) and Marston et al. (2020), respectively. Four hyperspectral measurements per cage were collected similarly to Marston et al. (2020), on 9 July, 7 August and 14 August of 2019; on 10 July, 15 July, 30 July, 21 August, 28 August and 4 September of 2020; and on 29 July, 3 August, 13 August, 17 August, 20 August, 22 August, 30 August and 10 September of 2021.

Hyperspectral measures were processed using the software ViewSpec Pro version 6.2.0 (ASD Inc., Boulder, CO, USA), and then averaged for each cage for each site-date. The averaged hyperspectral data were normalized using the following equation Marston et al. (2022):

$$NR\lambda_{pd} = \frac{R\lambda_{pd} \times R\lambda_{u}}{R\lambda_{ud}}$$

where $NR\lambda_{pd}$ is the normalized average hyperspectral reflectance at wavelength λ for plot p on date d, $R\lambda_{pd}$ is the average hyperspectral reflectance at wavelength λ for plot p on date d, $R\lambda_{u}$ is the average hyperspectral reflectance at wavelength λ for all plots u with less than 60 aphids per plant across all site-dates, and $R\lambda_{ud}$ is the average hyperspectral reflectance at wavelength λ for all plots u with less than 60 aphids per plant on date d. An average aphid density of less than 60 aphids per plant was used for the normalization because such SBA densities are unlikely to have adverse effects on soybean spectral reflectance (Alves et al., 2015; Marston et al., 2020).

2.1.3 Simulation of Sentinel-2 satellite spectral reflectance

The normalized ground-based hyperspectral reflectance from the cage studies (described in section 2.1.2) was used to simulate spectral reflectance of Sentinel-2 bands using the following equation (D’Odorico et al., 2013):

$$R(\omega) = \frac{\int_{\omega_{min}}^{\omega_{max}} R_h(\omega_i) \times SRF(\omega_i) d(\omega_i)}{\int_{\omega_{min}}^{\omega_{max}} SRF(\omega_i) d(\omega_i)}$$

where $R(\omega)$ is the simulated spectral reflectance of a Sentinel-2 band ω, $R_h(\omega_i)$ is the hyperspectral reflectance of the narrowbands ω_i measured at the ground level that correspond to the spectral response function *(SRF)* of the Sentinel-2 sensor for the band ω. *SRF* was calculated for each band ω using the following equation:

$$SRF(\omega_i) = \frac{SRF_A(\omega_i) + SRF_B(\omega_i)}{2}$$

where $SRF_A(\omega_i)$ and $SRF_B(\omega_i)$ are the spectral responses of the multispectral instrument of the Sentinel-2A and Sentinel-2B satellites, respectively, for the narrowbands ω_i present on both instruments. VIs used in previous studies assessing the relationship between plant spectral reflectance and different stressors (e.g., insect feeding and diseases) were also calculated using the simulated Sentinel-2 bands (Table 2) and used in the analyses.

Table 2. Selected vegetation indices for satellite-based assessment of soybean aphid in soybean

Index	Equation	Developed by	Implemented by	Stressor / Crop
Normalized Difference Vegetation Index	$NDVI = \frac{(B8 - B4)}{(B8 + B4)}$	Rouse et al. (1973)	(Yang et al., 2009)	Aphid / Wheat
			Chemura et al. (2017)	Leaf rust / Coffee
Green Normalized Difference Vegetation Index	$GNDVI = \frac{(B8 - B3)}{(B8 + B3)}$	Gitelson et al. (1996)	Chemura et al. (2017)	Leaf rust / Coffee
			(Reisig and Godfrey, 2006)	Aphid and mite / Cotton
Normalized Difference Red Edge Index	$NDRE = \frac{(B7 - B5)}{(B7 + B5)}$	Gitelson and Merzlyak (1994)	Liu et al. (2018)	Heavy metal / Rice
			Chemura et al. (2017)	Leaf rust / Coffee
Soil Adjusted Vegetation Index	$SAVI = 1.5 \times \frac{(B8A - B4)}{(B8A + B4 + 0.5)}$	Huete (1988)	Hawryło et al. (2018)	Bark beetle / Pine
			Yang et al. (2009)	Aphid / Wheat
Optimized Soil-Adjusted Vegetation Index	$OSAVI = 1.16 \times \frac{(B8A - B4)}{(B8A + B4 + 0.16)}$	Rondeaux et al. (1996)	Yang et al. (2009)	Aphid / Wheat
			(Reisig and Godfrey, 2006)	Aphid and mite / Cotton

2.2 Actual satellite measurements from commercial fields

2.2.1 Field-scale samples and data selection

From 2017 to 2019, a total of 107 commercial soybean fields were sampled from the V5 to R6 growth stages (Fehr and Caviness, 1977) in Minnesota, United States. Fields with soybean plants during earlier and later developmental stages were not included to avoid the effects of bare ground soil before soybean canopy closure and of physiological changes associated with plant maturity, respectively. On each sample date for each field, a representative number of soybean plants (around 40 plants) were randomly selected from throughout the field and visually inspected to estimate the abundance of SBA (Hodgson et al., 2004; Ragsdale et al., 2007). SBA abundance was estimated in the field using visual whole-plant counts immediately after pulling the selected plants from the ground (i.e., destructive sampling). Global positioning system coordinates were recorded for each field.

For commercial fields sampled more than once within a 7-day period, only one sample date with the highest average SBA density was selected. The time frame of 7 days was chosen based on the revisiting time of the Sentinel-2 satellites (Drusch et al., 2012). For each field with average SBA density above 60 aphids per plant, a corresponding field within 5 km sampled within 7 days, and with a density lower than 60 aphids per plant was selected to account for possible variability in time and space. The threshold of 60 aphids per plant was used for the same reasons described in section 2.1.2. Finally, field dates covered with clouds or with cloud shadows were excluded (see section 2.2.2 for more details), resulting in a total of 22 field dates for the statistical analyses. To ensure plant stage in these 22 fields would not be a confounding effect in the subsequent analyses, Pearson's correlation between plant stage and the average number of aphids per plant was performed (R package, *function*: stats, *cor.test*; R Core Team, 2021) and this effect was not significant ($r = 0.17$, $t = 0.80$, $df = 20$, p value = 0.437).

2.2.2 Satellite imagery acquisition and data processing

Multispectral Sentinel-2 satellite level 1C (top of atmosphere reflectance) imagery were downloaded from the European Space Agency Copernicus Open Access Hub data repository (ESA, 2023). Each multispectral image was visually inspected for the presence of clouds using the preview option on the Copernicus website, and only images with the following criteria were downloaded: 1) image acquired within 7 days of field sampling; and 2) less than 20% clouds, or less than 40% clouds as long as clouds were confined to one side of the image (opposite to sampled fields).

Level 1C imagery were atmospheric-, terrain- and cirrus-corrected and converted to level 2A (bottom of atmosphere reflectance, in digital numbers) imagery with 20-m resolution using the standalone sen2cor processor (Main-Knorn et al., 2017) via Windows prompt command. Sen2cor version 2.5.5 and version 2.10.1 were used for imagery from 2017 and from 2018 to 2019, respectively, because files previous to 2018 cannot be processed with new versions of sen2cor due to a change in the metadata structure of the imagery files implemented after 2017.

Selected satellite level 2A imagery with 20-m resolution was processed and boundaries of the commercial soybean fields were delineated in ArcMap version 10.8.2 (ESRI, 2021). Surface reflectance and VIs were calculated using the "raster calculation" tool in ArcMap. Surface reflectance was obtained for each band of each multispectral image by dividing the digital number of each pixel by 10,000 (Main-Knorn et al., 2017). VIs were obtained using the equations described in Table 2. Bare ground areas result in values of NDVI < 0.4 (Zhang et al., 2015). Thus, bare ground pixels were removed from all images using NDVI < 0.4 as a reference.

Each field boundary was manually delineated using its respective true color composite image (i.e., colored image resulting from the satellite's red, green and blue color channels) as a visual reference, and fields covered by clouds or cloud shade were excluded. Pixels within 20 m of the field edge (i.e., field boundary) were excluded using the “buffer” tool in ArcMap to avoid the influence of surrounding areas on field spectral reflectance. Finally, the average reflectance of each field was calculated for all bands and VIs of their respective Sentinel-2 multispectral image using the “zonal statistics as table” tool in ArcMap.

2.3 Statistical analyses

The software R version 4.1.2 (R Core Team, 2021) and RStudio Desktop version 2021.9.2.382 (RStudio Team, 2021) were used to perform all analyses and to create graphs. CAD from cage studies and average number of aphids per plant from commercial soybean fields were log-transformed as $ln(X + 1)$, where X corresponds to CAD from each cage or the average number of aphids per plant per field. Simulated and actual Sentinel-2 bands 1, 8, 9, 10, 11 and 12 were not included in this study because they: i) have low spatial resolution (i.e., bands 1, 9, and 10 > 20 m), ii) offer redundant information (i.e., bands 8 and 8A), or iii) use wavelengths outside the detection range of the hyperspectral spectroradiometer used in the cage studies between 2017 and 2021 (i.e., bands 10, 11, and 12 >1000 μm). Thus, only simulated and actual Sentinel-2 bands 2, 3, 4, 5, 6, 7, and 8A were used in this study.

For the simulation of satellite measurements using caged plots, the effects of log-transformed CAD on simulated Sentinel-2 bands and VIs were analyzed using random intercept linear mixed models with date nested in year as a random factor (lme4, *lmer*; Bates et al., 2015). Degrees of freedom and p values were estimated for each model using the Satterthwaite method (lmer, *anova*; Kuznetsova et al., 2017). Model assumptions (linearity, normality of residuals, normality of random effects, and homogeneity of variance) were visually checked with diagnostic plots (performance, *check_model*; Lüdecke et al., 2021). Conditional and marginal R^2 values were obtained using the Nakagawa's R^2 for mixed models (performance, *r2*; Lüdecke et al., 2021).

For the actual satellite measurements from commercial fields, the effects of log-transformed average number of aphids per field on average Sentinel-2 spectral reflectance and VIs of soybean fields were analyzed using general linear models (stats, *lm*; R Core Team, 2021). Model assumptions (linearity, normality of residuals, and homogeneity of variance) were visually checked as described above.

For classification of SBA infestations in commercial fields as above or below the economic threshold of 250 aphids per plant, LSVM models were developed using simulated Sentinel-2 bands and VIs from the caged plots, and were tested on actual Sentinel-2 data from commercial soybean fields. Initially, recursive feature elimination using 10-fold repeated cross-validation with 3 repetitions (caret, *rfe*; Kuhn, 2008) was used to select the best predictors with highest accuracy to be used in the LSVM models.

Then, models containing combinations of 1, 2, 3 or 4 of the selected predictors (i.e., simulated Sentinel-2 bands 7 and 8A, and simulated Sentinel-2-based VIs SAVI and OSAVI) were further fine-tuned (caret, *rfe*; Kuhn, 2008). For each model, fine-tuning was done using 10-fold repeated cross-validation with 3 repetitions, a grid-based search between 0.01 and 1000 for the parameter C, and weights to each class (i.e., above and below the economic threshold) as a proportion of the total number of samples in each class to account for class imbalance. Fine-tuned models containing 2 predictors had higher accuracy and therefore were used in the final models. Final models with combinations of 2 of the selected predictors were obtained (caret, *train*; Kuhn, 2008) and their overall accuracy and Cohen's kappa values were compared using pairwise Bonferroni-corrected t-tests (caret, *resamples* followed by *diff*; Kuhn, 2008). Final models were tested (stats, *predict*; R Core Team, 2021) on actual Sentinel-2 data from commercial soybean fields infested with SBA, and model classification metrics were obtained using confusion matrices (caret, *confusionMatrix*; Kuhn, 2008). Similar to Marston et al. (2022), the final model was selected based on overall highest accuracy, Cohen's kappa, sensitivity and specificity. Cohen's kappa measures observed accuracy considering the expected accuracy that might occur by random chance, sensitivity measures true positive classification (i.e., correctly classifying commercial soybean fields above the economic threshold), and specificity measures true negative classification (i.e., correctly classifying commercial soybean fields below the economic threshold) (Allouche et al., 2006; Marston et al., 2022).

3 Results

In the simulation of satellite measurements using caged plots, increases in CAD were significantly associated with a reduction in the spectral reflectance of simulated Sentinel-2 bands 5, 6, 7 and 8A, and of the simulated Sentinel-2-based VIs NDVI, GNDVI, SAVI, OSAVI and NDRE (p values of slopes < 0.05) from caged soybean plants (Table 3). Slope values from significant regressions ranged from -5.8×10^{-4} to -9.77×10^{-3} (Table 3).

A similar response was observed for the actual satellite measurements from commercial soybean fields, where an increase in the average number of aphids per plant per field was significantly associated with a reduction (p values of slopes < 0.05) in the spectral reflectance of actual Sentinel-2 bands 6, 7 and 8A, as well as the Sentinel-2-based VIs NDVI, SAVI and OSAVI (Table 4). Slopes of significant regressions ranged from -5.43×10^{-3} to -1.81×10^{-2} (Table 4). The linear regressions and actual spectral reflectance of sampled soybean fields are represented in Figure 1 for the significant Sentinel-2 bands and Sentinel-2-based VIs.

Table 3. Summary outputs, analysis of variance using the Satterthwaite's method, and Nakagawa's R^2 values (conditional and marginal) of linear mixed models estimating the effects of log-transformed cumulative aphid days for soybean aphid on simulated Sentinel-2 satellite bands and vegetation indices from ground-based hyperspectral data of cage studies done in 2013 and 2014, and from 2017 to 2021 in Minnesota, United States, and in 2017 and 2018 in Iowa, United States

Model	Intercept ± SE	Slope ± SE	F	df	p value†	Conditional $R^{2‡}$	Marginal $R^{2‡}$
2	$2.91 \times 10^{-2} \pm 3.35 \times 10^{-3}$	$-1.35 \times 10^{-4} \pm 1.17 \times 10^{-4}$	1.32	1;536.5	0.250	0.846	0.001
3	$5.66 \times 10^{-2} \pm 6.14 \times 10^{-3}$	$-3.52 \times 10^{-4} \pm 2.01 \times 10^{-4}$	3.08	1;535.5	0.080	0.870	0.001
4	$2.79 \times 10^{-2} \pm 3.93 \times 10^{-3}$	$-1.37 \times 10^{-4} \pm 1.37 \times 10^{-4}$	1.01	1;536.1	0.316	0.857	0.000
5	$8.39 \times 10^{-2} \pm 8.37 \times 10^{-3}$	$-5.80 \times 10^{-4} \pm 2.76 \times 10^{-4}$	4.42	1;535.4	**0.036**	0.874	0.002
6	$4.22 \times 10^{-1} \pm 1.27 \times 10^{-2}$	$-6.04 \times 10^{-3} \pm 8.83 \times 10^{-4}$	46.77	1;543.0	**<0.001**	0.688	0.044
7	$5.65 \times 10^{-1} \pm 1.53 \times 10^{-2}$	$-9.77 \times 10^{-3} \pm 1.16 \times 10^{-3}$	70.77	1;550.4	**<0.001**	0.684	0.068
8A	$5.82 \times 10^{-1} \pm 1.58 \times 10^{-2}$	$-9.60 \times 10^{-3} \pm 1.17 \times 10^{-3}$	67.7	1;549.4	**<0.001**	0.697	0.062
NDVI	$9.08 \times 10^{-1} \pm 1.17 \times 10^{-2}$	$-1.85 \times 10^{-3} \pm 4.87 \times 10^{-4}$	14.43	1;537.2	**<0.001**	0.839	0.007
GNDVI	$8.21 \times 10^{-1} \pm 1.59 \times 10^{-2}$	$-2.64 \times 10^{-3} \pm 5.85 \times 10^{-4}$	20.46	1;535.9	**<0.001**	0.867	0.008
SAVI	$7.43 \times 10^{-1} \pm 1.20 \times 10^{-2}$	$-8.48 \times 10^{-3} \pm 9.19 \times 10^{-4}$	85.15	1;550.6	**<0.001**	0.686	0.081
OSAVI	$8.32 \times 10^{-1} \pm 1.03 \times 10^{-2}$	$-6.01 \times 10^{-3} \pm 6.90 \times 10^{-4}$	75.78	1;546.3	**<0.001**	0.739	0.060
NDRE	$7.43 \times 10^{-1} \pm 1.93 \times 10^{-2}$	$-3.58 \times 10^{-3} \pm 7.38 \times 10^{-4}$	23.47	1;536.1	**<0.001**	0.865	0.010

† Significant p values (< 0.05) are boldfaced

‡ Conditional R^2 refers to the variance explained by both fixed and random factors, and marginal R^2 refers to the variance explained by fixed factors only

Table 4. Summary outputs, analysis of variance, and R^2 values of general linear models estimating the effects of log-transformed average number of soybean aphids per plant on actual Sentinel-2 satellite bands and vegetation indices from commercial soybean fields sampled from 2017 to 2019 in Minnesota, United States

Model	Intercept ± SE	Slope ± SE	F	df	p value†	Multiple $R^{2‡}$	Adjusted $R^{2‡}$
2	$2.34 \times 10^{-2} \pm 3.10 \times 10^{-3}$	$1.04 \times 10^{-3} \pm 6.51 \times 10^{-4}$	2.53	1;20	0.127	0.112	0.068
3	$4.55 \times 10^{-2} \pm 4.91 \times 10^{-3}$	$2.35 \times 10^{-4} \pm 1.03 \times 10^{-3}$	0.05	1;20	0.822	0.003	-0.047
4	$2.20 \times 10^{-2} \pm 2.89 \times 10^{-3}$	$6.23 \times 10^{-4} \pm 6.06 \times 10^{-4}$	1.06	1;20	0.316	0.050	0.003
5	$7.28 \times 10^{-2} \pm 8.04 \times 10^{-3}$	$-2.38 \times 10^{-4} \pm 1.69 \times 10^{-3}$	0.02	1;20	0.889	0.001	-0.049
6	$4.32 \times 10^{-1} \pm 2.53 \times 10^{-2}$	$-1.25 \times 10^{-2} \pm 5.32 \times 10^{-3}$	5.49	1;20	**0.030**	0.215	0.176
7	$5.94 \times 10^{-1} \pm 3.29 \times 10^{-2}$	$-1.78 \times 10^{-2} \pm 6.91 \times 10^{-3}$	6.61	1;20	**0.018**	0.248	0.211
8A	$6.20 \times 10^{-1} \pm 3.46 \times 10^{-2}$	$-1.81 \times 10^{-2} \pm 7.25 \times 10^{-3}$	6.20	1;20	**0.022**	0.237	0.198
NDVI	$9.34 \times 10^{-1} \pm 1.22 \times 10^{-2}$	$-5.43 \times 10^{-3} \pm 2.56 \times 10^{-3}$	4.48	1;20	**0.047**	0.183	0.142
GNDVI	$8.66 \times 10^{-1} \pm 1.77 \times 10^{-2}$	$-6.13 \times 10^{-3} \pm 3.71 \times 10^{-3}$	2.74	1;20	0.114	0.120	0.076
SAVI	$7.88 \times 10^{-1} \pm 2.69 \times 10^{-2}$	$-1.49 \times 10^{-2} \pm 5.65 \times 10^{-3}$	7.01	1;20	**0.015**	0.259	0.222
OSAVI	$8.68 \times 10^{-1} \pm 1.93 \times 10^{-2}$	$-1.07 \times 10^{-2} \pm 4.06 \times 10^{-3}$	6.97	1;20	**0.016**	0.258	0.221
NDRE	$7.86 \times 10^{-1} \pm 2.56 \times 10^{-2}$	$-7.24 \times 10^{-3} \pm 5.38 \times 10^{-3}$	1.82	1;20	0.193	0.083	0.038

† Significant p values (< 0.05) are boldfaced

‡ Multiple R^2 refers to the variance explained by fixed factors, and adjusted R^2 refers to the variance explained by fixed factors adjusted by the number of predictors in the model

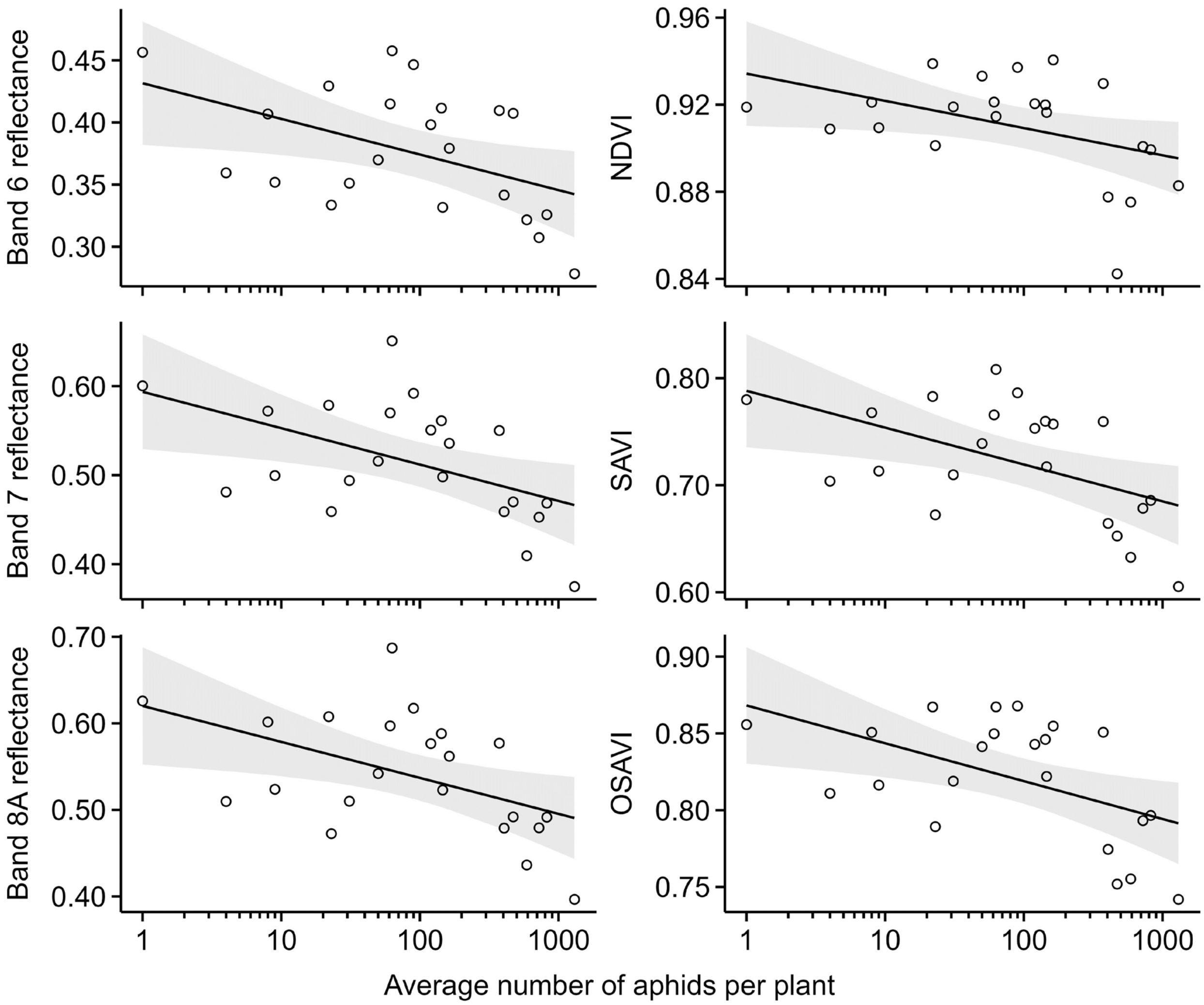

Fig. 1. Linear regressions and 95% confidence bands representing significant effects of soybean aphid infestations on actual Sentinel-2 satellite bands (6, 7 and 8A) and Sentinel-2-based vegetation indices (NDVI, SAVI and OSAVI) from commercial soybean fields sampled from 2017 to 2019 in Minnesota, United States.

Four LSVM models were able to classify SBA infestations in soybean fields as above or below the economic threshold of 250 aphids per plant, using actual Sentinel-2 individual band spectral reflectance and Sentinel-2-based VIs, with a significant improvement (p values < 0.05) over the no-information rate (Table 5). Model 2 had numerically higher accuracy (91%) and Cohen's kappa (79%), but Pairwise Bonferroni-corrected t-tests indicated no significant differences (p values > 0.05) among the four LSVM models (Table 6). The specificity (i.e., correctly classifying fields below the economic threshold) of models 1 and 2 was the same (93.3%) but numerically lower than models 3 and 4 (100%). However, the sensitivity (i.e., correctly classifying fields above the economic threshold) and balanced accuracy (85.7 and 89.5%, respectively) of model 2 were also numerically the highest. Thus, model 2, using actual Sentinel-2 satellite spectral reflectance from band 7 and the Sentinel-2-based SAVI, was chosen for the classification of SBA infestations in soybean fields.

The average number of aphids per plant and classification outcomes using the optimal LSVM model (i.e., model 2) for the commercial soybean fields are represented in Figure 2. SBA infestations were above the economic threshold of 250 aphids per plant in fields 1 through 7, with average SBA densities ranging from 373 to 1303 aphids per plant. These fields were correctly classified as above the economic threshold, except for field 7, which is closest to the threshold (Fig. 2). SBA infestations in fields 8 through 22 were below the economic threshold, with average SBA densities ranging from 0 to 162 aphids per plant. These fields were correctly classified as below the economic threshold, except for field 17 (Fig. 2). Field locations and corresponding classification outcomes are represented in Figure 3. The spectral reflectance of actual Sentinel-2 satellite band 7 and the Sentinel-2-based SAVI (i.e., used in the selected SVM model) are represented in Figure 4 for two soybean fields with high and low SBA infestations.

Table 5. Training and testing performance statistics of significant linear support vector machine models using 2 predictors (Input) for the classification of commercial soybean fields infested with soybean aphids as above (positive class) or below (negative class) an economic threshold of 250 aphids per plant. Models were trained on simulated Sentinel-2 satellite bands and vegetation indices (VIs) from ground-based hyperspectral data of cage studies done in 2013 and 2014, and from 2017 to 2021 in Minnesota, United States, and in 2017 and 2018 in Iowa, United States. Models were tested on actual Sentinel-2 satellite bands and vegetation indices from commercial soybean fields sampled from 2017 to 2019 in Minnesota, United States.

	Model 1	Model 2	Model 3	Model 4
Training using simulated satellite data				
Input (band and/or VI)	8A, SAVI	7, SAVI	7, OSAVI	SAVI, OSAVI
Parameter C	0.100	0.750	1.000	0.750
Accuracy	0.864	0.860	0.861	0.860
Cohen's kappa	0.499	0.487	0.490	0.467
Testing using actual satellite data				
Input (band and/or VI)	8A, SAVI	7, SAVI	7, OSAVI	SAVI, OSAVI
Accuracy	0.864	0.910	0.864	0.864
95% confidence interval	0.651–0.971	0.708–0.989	0.651–0.971	0.651–0.971
No-information rate (NIR)	0.682	0.682	0.682	0.682
p value (accuracy > NIR)	0.048	0.013	0.048	0.048
Cohen's kappa	0.673	0.790	0.645	0.645
Sensitivity	0.714	0.857	0.571	0.571
Specificity	0.933	0.933	1.000	1.000
Balanced accuracy	0.824	0.895	0.786	0.786

Table 6. Pairwise Bonferroni-corrected t-tests of accuracy and Cohen's kappa from significant linear support vector machine models using 2 predictors for the classification of soybean fields infested with soybean aphids as above (positive class) or below (negative class) an economic threshold of 250 aphids per plant. Predictors (Input) are actual Sentinel-2 satellite bands and vegetation indices (VIs) from commercial soybean fields sampled from 2017 to 2019 in Minnesota, United States.

	Model 1	Model 2	Model 3	Model 4
Input (band and/or VI)	8A, SAVI	7, SAVI	7, OSAVI	SAVI, OSAVI
		Accuracy		
Model 1		-5.22×10^{-03}	-7.11×10^{-04}	-1.64×10^{-03}
Model 2	1		4.5×10^{-03}	3.58×10^{-03}
Model 3	1	1		-9.25×10^{-04}
Model 4	1	1	1	
		Cohen's kappa		
Model 1		-1.88×10^{-02}	3.54×10^{-02}	4.98×10^{-02}
Model 2	1		5.42×10^{-02}	6.86×10^{-02}
Model 3	1	1		1.44×10^{-02}
Model 4	1	0.818	1	

Within each matrix, the upper diagonal indicates the differences between model accuracies or Cohen's kappa, and the lower diagonal indicates the corresponding Bonferroni-corrected p values.

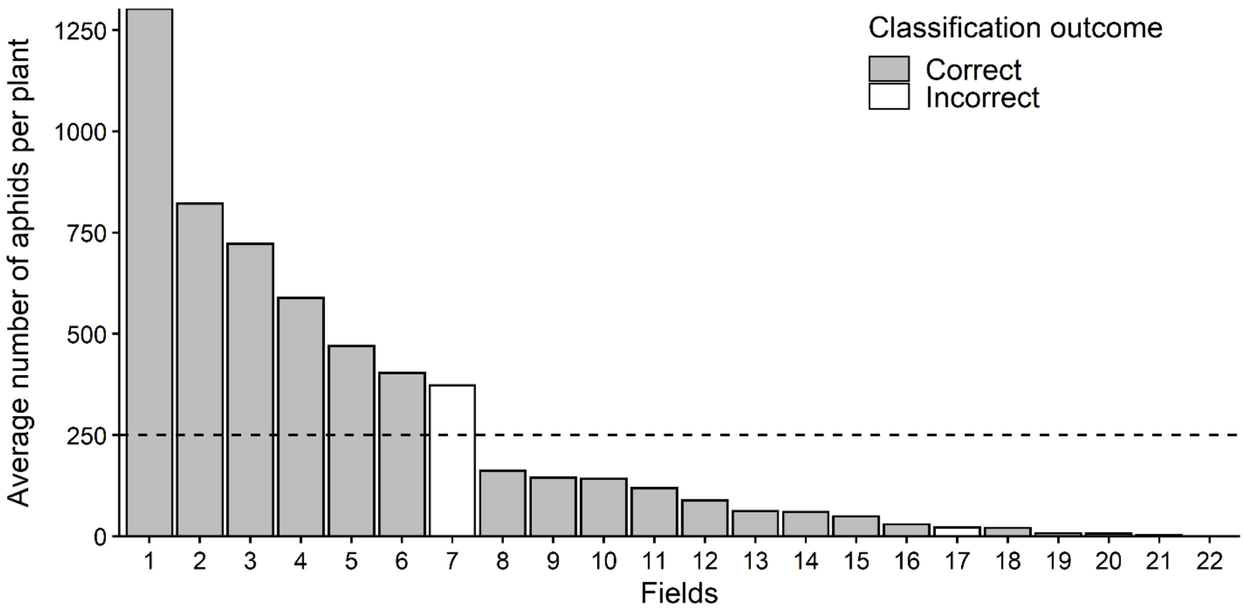

Fig. 2. Average number of aphids per plant from commercial soybean fields sampled from 2017 to 2019 in Minnesota, United States. Fields were classified by the best linear support vector machine model as above or below an economic threshold of 250 aphids per plant (horizontal dashed line) using actual Sentinel-2 satellite spectral reflectance from band 7 and the Sentinel-2-based Soil Adjusted Vegetation Index (SAVI).

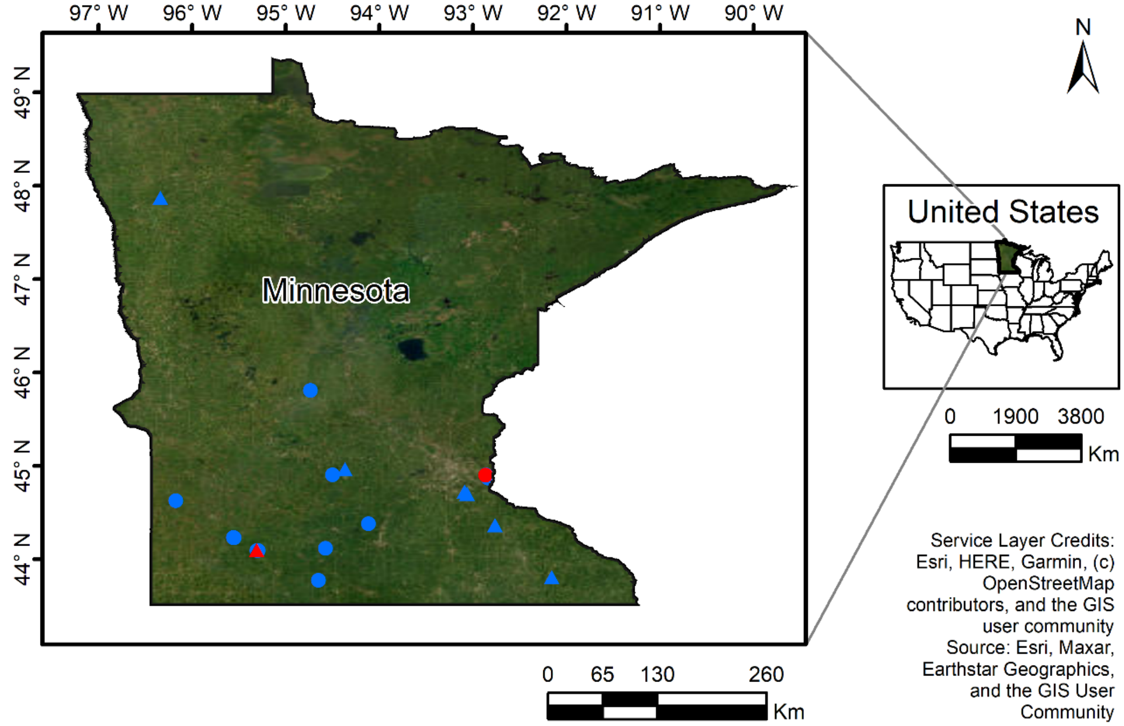

Fig. 3. Commercial soybean fields (n = 22) sampled from 2017 to 2019 in Minnesota, United States, for the presence of soybean aphids. Circles and triangles represent soybean fields below and above an economic threshold of 250 aphids per plant, respectively. Blue and red symbols indicate soybean fields correctly and incorrectly classified, respectively, by the best linear support vector machine model using actual Sentinel-2 satellite spectral reflectance from band 7 and the Sentinel-2-based Soil Adjusted Vegetation Index (SAVI). Note that some of the symbols are completely overlapped.

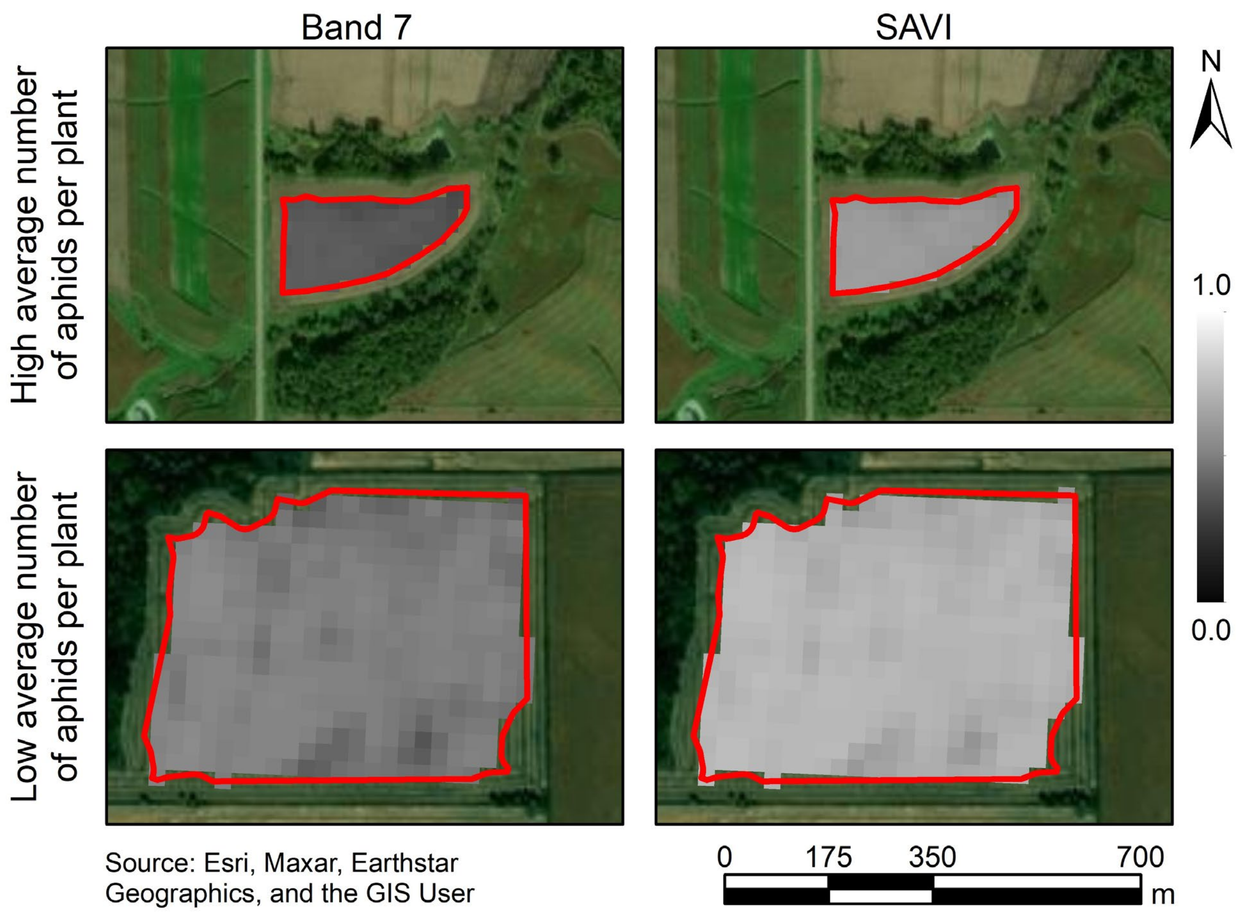

Fig. 4. Soybean fields with high and low average number of aphids per plant. Field boundaries are indicated by a red line 20 m from the edge. Within field boundaries, the spectral reflectance imagery with 20-m resolution of actual Sentinel-2 satellite band 7 and the Sentinel-2-based Soil Adjusted Vegetation Index (SAVI) are represented in grayscale.

4 Discussion

This is the first study to show that satellite-based data can be used for sensing of plant stress associated with SBA infestations. Both the simulation of satellite spectral measurements from small plots and the retrospective validation from commercial soybean fields using actual Sentinel-2 imagery indicated a significant relationship between SBA and the spectral reflectance of soybean plants. Furthermore, four LSVM models for the classification of SBA infestations in soybean fields were successfully developed.

Insect injury can cause physicochemical changes to plants, which have direct and indirect effects on canopy spectral reflectance, especially in the visible and near-infrared regions (Abd El-Ghany et al., 2020; Rhodes et al., 2022). Changes in plant reflectance in the visible and near-infrared region due to injury of leaf-feeding insects can happen simultaneously and they are usually associated, among other things, with a reduction in leaf chlorophyll content and changes in the leaf structure, respectively (Jackson, 1986; Abd El-Ghany et al., 2020). These effects have been described for multiple pests, including aphids (Reisig and Godfrey, 2006; Yang et al., 2009; Luo et al., 2013). However, SBA injury has been found to mainly cause a decrease in the near-infrared spectral reflectance of soybean plants (Alves et al., 2015; Marston et al., 2020). This lack of response in the visible region for SBA can be because feeding by this insect might not affect the chlorophyll content of soybean leaves (Macedo et al., 2003; Alves et al., 2015). Nevertheless, significant effects on normalized difference vegetation index (NDVI), a vegetation index incorporating both visible and near-infrared bands, have also been detected for SBA (Alves et al., 2015; Marston et al., 2020) and other aphids (Reisig and Godfrey, 2006; Yang et al., 2009; Elliott et al., 2015).

The present study, using simulated and actual satellite measurements, showed effects of SBA on plant reflectance in the near-infrared region, which is similar to previous findings for this pest using proximal remote sensing (Alves et al., 2015; Marston et al., 2020). Simulated satellite spectral reflectance indicated four bands in the near-infrared region and five VIs with the potential for detecting plant stress caused by SBA feeding. In addition, three Sentinel-2 bands in the near-infrared region and three VIs were confirmed to be sensitive to SBA infestations in soybean fields by the retrospective validation using actual Sentinel-2 imagery. The simulation of satellite bands and VIs, including Sentinel-2, from ground-based hyperspectral reflectance data has also been used by other studies (Martins et al., 2017; Osco et al., 2019). These findings corroborate the possibility of using simulation of satellite spectral reflectance for the screening of potential satellite bands and VIs for the detection of plant stressors in the field. The collection of ground truth data for studies using satellite imagery is time-consuming, so the early detection of potential satellite bands and VIs from small semi-field conditions could save time in future studies for other pests before effort is put into surveilling large areas.

Similar to previous work documenting the negative effects of SBA on soybean NDVI (Alves et al., 2015; Marston et al., 2020), both the simulated and actual Sentinel-2-based NDVI had a significant decrease in this study, as well as soil-adjusted vegetation index (SAVI) and optimized soil-adjusted vegetation index (OSAVI). These indices include bands in both visible and near-infrared regions, but SAVI and OSAVI also incorporate a correction factor to account for the influence of soil background (Huete, 1988; Rondeaux et al., 1996). Under real field conditions, the effects of soil are of utmost importance because the spectral reflectance of plant canopy is the result of the combination of all sources of reflectance in the field (Rondeaux et al., 1996; Mulla, 2013). Thus, it is expected that soil background can have an effect on the detection and classification of plant stresses in the field. In fact, the optimal LSVM model for the classification of SBA infestations in soybean fields in this study was developed using the Sentinel-2 band 7 (near-infrared) and SAVI.

The LSVM model selected in this study was able to classify SBA infestations as above or below the economic threshold of 250 aphids per plant with an accuracy of 91%. The ability to classify fields into actionable classes has direct implications on decision-making for the management of pests because misclassifications can lead to treating fields unnecessarily or to economic losses if highly infested fields are not treated (Reay-Jones et al., 2009). The latter is more critical and error rates ≤ 10% are desirable for the classification of insect pest infestations in the field (Hodgson et al., 2004; Reay-Jones et al., 2009). Out of the 22 commercial soybean fields used to validate the LSVM model developed in this study, only two fields (i.e., one above and one below the economic threshold) were misclassified. Furthermore, all the fields with SBA infestations above the economic injury level of 674 aphids per plant were correctly classified. The accuracy of a variety of machine learning models developed by previous studies for the classification of diseases or insect pests using Sentinel-2 imagery ranged from 71 to 89% for wheat (Yuan et al., 2014, 2017) and from 67 to 96% for forests (Hawryło et al., 2018; Abdullah et al., 2019). Hence, the LSVM model developed in this study is accurate and it has the potential to be incorporated into monitoring programs for SBA in soybean.

Remote sensing using satellites, like any other remote sensing platform, is not free from limitations (e.g., cloud cover, time lag for delivery of management recommendations, and multiple stressors). Specifically for SBA in soybean, the established economic threshold for this pest provides a seven-day lead time before infestations reach the economic injury level (Ragsdale et al. 2007). Considering the revisit frequency of 3–5 days of the Sentinel-2 satellites, and assuming cloud-free days not being a limitation, it is likely that SBA infestations could be detected prior to reaching the economic injury level. The presence of clouds can limit the availability of satellite imagery and, therefore, the remote sensing of plant stresses caused by pests (Mulla, 2013; Rhodes et al., 2022). Despite the high revisit frequency (3–5 days) of the Sentinel-2 satellites, the occurrence of cloudy days at critical periods during the growing season may restrain the use of tools such as the one developed in this study, and compromise the delivery of timely management recommendations. However, processing of satellite data for the generation of management recommendations for SBA must be done quickly (i.e., ideally within the same day of acquisition of the remote sensing imagery) to allow enough time for farmers to make arrangements to treat infested soybean fields whenever infestations reach the economic threshold.

Multiple stressors (e.g., pests and diseases) can occur simultaneously under field conditions. Changes to the spectral reflectance of soybean in the visible and/or near infrared have been documented for other pests (Iost Filho et al., 2022; Ribeiro et al., 2022). The LSVM model developed in this study is robust because it was validated using data from real field conditions from multiple years, possibly with the presence of other stressors in the soybean fields with SBA. However, the occurrence of other stressors was not documented, so such impacts on the results presented here cannot be evaluated. Thus, additional studies investigating possible confounding effects of multiple stressors to soybean spectral reflectance are encouraged to further refine remote sensing technologies for soybean integrated pest management.

5 Conclusions

The classification of fields using satellite data could be used to prioritize fields for more intensive ground- or drone-based scouting, or to directly inform decision-making for individual fields. Satellites such as Sentinel-2 cover large areas, and the increase in spatial and temporal resolutions of satellite imagery observed in the last few years might facilitate actionable use of satellite data with greater efficiency and reduced costs to scout fields for pests (Drusch et al., 2012; Frampton et al., 2013; Rhodes et al., 2022). Field treatment decisions for SBA are still based on average field-level density estimates performed at one-week intervals (Koch et al., 2016). Therefore, Sentinel-2 with spectral data collected every 3–5 days with a 10–20 m resolution appear to be sufficient for field-level decision-making. The developed LSVM model can be used to assist regional monitoring and field-level scouting for this pest. Thus, the findings of this study will help to further advance regional and local management programs for SBA, and guide future studies on the use of satellite imagery for other pests.

Acknowledgments

We thank James Menger, Dr. Wally Rich, Dr. Anh Tran, Dr. Anthony Hanson, Kathryn Pawley, Annika Asp, Celia Silverstein, Mads Bartz, Kendra Moran, Courtney Garrison Hickey, Alissa Geske, Pheylan Anderson, Claire Lotzer, Traci Eicholz, Narayan Bhagroo, Julia Struatman, Dr. Daniela Pezzini, Obiratanea de Silva Queiroz, Rafael Carlesso Aita, Dr. Nádia Bueno, Gregory VanNostrand, Ashley Dean, and Kimon Karelis, Dr. Rosa (Tina) Lozano, Rafael Carlesso Aita, Grace Doyle, Samantha

Janecek, Gunnar Morris, Ellen Adjeiwaa, and Gloria Melotto for assistance in collecting field data and/or managing research plots.

Funding

This project was supported by the Minnesota Invasive Terrestrial Plants and Pests Center through the Minnesota Environment and Natural Resources Trust Fund. Some of the historical data used in the analyses were funded by the Minnesota Soybean Research and Promotion Council, the Agriculture and Food Research Initiative of the USDA’s National Institute of Food and Agriculture (competitive grant 2016-70006-25828), the University of Minnesota MnDRIVE Initiative, and National Council for Scientific and Technological Development (CNPq/Brazil).

References

Abd El-Ghany, N.M., Abd El-Aziz, S.E., Marei, S.S., 2020. A review: application of remote sensing as a promising strategy for insect pests and diseases management. Environ. Sci. Pollut. Res. 27, 33503–33515. https://doi.org/10.1007/s11356-020-09517-2

Abdullah, H., Skidmore, A.K., Darvishzadeh, R., Heurich, M., 2019. Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (*Ips typographus*, L.) compared with Landsat-8. Remote Sens. Ecol. Conserv. 5, 87–106. https://doi.org/10.1002/rse2.93

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Alves, T.M., Macrae, I. V., Koch, R.L., 2015. Soybean aphid (Hemiptera: Aphididae) affects soybean spectral reflectance. J. Econ. Entomol. 108, 2655–2664. https://doi.org/10.1093/jee/tov250

Alves, T.M., Moon, R.D., MacRae, I. V., Koch, R.L., 2019. Optimizing band selection for spectral detection of *Aphis glycines* Matsumura in soybean. Pest Manag. Sci. 75, 942–949. https://doi.org/10.1002/ps.5198

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67. https://doi.org/10.18637/jss.v067.i01

Bueno, A.F., Panizzi, A.R., Hunt, T.E., Dourado, P.M., Pitta, R.M., Gonçalves, J., 2021. Challenges for adoption of integrated pest management (IPM): the soybean example. Neotrop. Entomol. 50, 5–20. https://doi.org/10.1007/s13744-020-00792-9

Cavaco, A.M., Utkin, A.B., Marques da Silva, J., Guerra, R., 2022. Making sense of light: the use of optical spectroscopy techniques in plant sciences and agriculture. Appl. Sci. 12, 997. https://doi.org/10.3390/app12030997

Chemura, A., Mutanga, O., Dube, T., 2017. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis. Agric. 18, 859–881. https://doi.org/10.1007/s11119-016-9495-0

D’Odorico, P., Gonsamo, A., Damm, A., Schaepman, M.E., 2013. Experimental evaluation of Sentinel-2 spectral response functions for NDVI time-series continuity. IEEE Trans. Geosci. Remote Sens. 51, 1336–1348. https://doi.org/10.1109/TGRS.2012.2235447

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Elliott, N.C., Backoulou, G.F., Brewer, M.J., Giles, K.L., 2015. NDVI to detect sugarcane aphid injury to grain sorghum. J. Econ. Entomol. 108, 1452–1455. https://doi.org/10.1093/jee/tov080

ESA (European Space Agency), 2023. Copernicus Open Access Hub [WWW Document]. URL https://scihub.copernicus.eu/ (accessed 5.31.23).

ESA (European Space Agency), 2022. Sentinel-2 Spectral Response Functions (S2-SRF) [WWW Document]. URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-

/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses (accessed 6.16.23).

ESA (European Space Agency), 2015. MultiSpectral Instrument (MSI) Overview [WWW Document]. URL https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument (accessed 6.16.23).

ESRI (Environmental Systems Research Institute) Inc., 2021. ArcGIS Desktop.

Fehr, W.R., Caviness, C.E., 1977. Stages of Soybean Development, Cooperative Extension Service. Ames.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. J. Plant. Physiol. 143, 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Hanafi, A., Radcliffe, E.B., Ragsdale, D.W., 1989. Spread and control of potato leafroll virus in Minnesota. J. Econ. Entomol. 82, 1201–1206. https://doi.org/10.1093/jee/82.4.1201

Hawryło, P., Bednarz, B., Wężyk, P., Szostak, M., 2018. Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. Eur. J. Remote Sens. 51, 194–204. https://doi.org/10.1080/22797254.2017.1417745

Hesler, L.S., Beckendorf, E.A., 2021. Soybean aphid infestation and crop yield in relation to cultivar, foliar insecticide, and insecticidal seed treatment in South Dakota. Phytoparasitica 49, 971–981. https://doi.org/10.1007/s12600-021-00914-y

Hodgson, E.W., Burkness, E.C., Hutchison, W.D., Ragsdale, D.W., 2004. Enumerative and binomial sequential sampling plans for soybean aphid (Homoptera: Aphididae) in soybean. J. Econ. Entomol. 97, 2127–2136. https://doi.org/10.1093/jee/97.6.2127

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Iost Filho, F.H., de Bastos Pazini, J., de Medeiros, A.D., Rosalen, D.L., Yamamoto, P.T., 2022. Assessment of injury by four major pests in soybean plants using hyperspectral proximal imaging. Agronomy 12, 1516. https://doi.org/10.3390/agronomy12071516

Jackson, R.D., 1986. Remote sensing of biotic and abiotic plant stress. Annu. Rev. Phytopathol. 24, 265–287. https://doi.org/10.1146/annurev.py.24.090186.001405

Koch, R.L., Potter, B.D., Glogoza, P.A., Hodgson, E.W., Krupke, C.H., Tooker, J.F., DiFonzo, C.D., Michel, A.P., Tilmon, K.J., Prochaska, T.J., Knodel, J.J., Wright, R.J., Hunt, T.E., Jensen, B., Varenhorst, A.J., McCornack, B.P., Estes, K.A., Spencer, J.L., 2016. Biology and economics of recommendations for insecticide-based management of soybean aphid. Plant Health Prog. 17, 265–269. https://doi.org/10.1094/PHP-RV-16-0061

Kuhn, M., 2008. Building predictive models in *R* using the caret package. J. Stat. Softw. 28. https://doi.org/10.18637/jss.v028.i05

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/jss.v082.i13

Liu, M., Wang, T., Skidmore, A.K., Liu, X., 2018. Heavy metal-induced stress in rice crops detected using multi-temporal Sentinel-2 satellite images. Sci. Total Environ. 637–638, 18–29. https://doi.org/10.1016/j.scitotenv.2018.04.415

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., Makowski, D., 2021. performance: an R package for assessment, comparison and testing of statistical models. J. Open Source Softw. 6, 3139. https://doi.org/10.21105/joss.03139

Luo, J., Huang, W., Zhao, J., Zhang, J., Zhao, C., Ma, R., 2013. Detecting aphid density of winter wheat leaf using hyperspectral measurements. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 690–698. https://doi.org/10.1109/JSTARS.2013.2248345

Ma, Y., Lu, J., Huang, X., 2023. Damage diagnosis of *Pinus yunnanensis* canopies attacked by Tomicus using UAV hyperspectral images. Forests 14, 61. https://doi.org/10.3390/f14010061

Macedo, T.B., Bastos, C.S., Higley, L.G., Ostlie, K.R., Madhavan, S., 2003. Photosynthetic responses of soybean to soybean aphid (Homoptera: Aphididae) injury. J. Econ. Entomol. 96, 188–193. https://doi.org/10.1093/jee/96.1.188

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2, in: Bruzzone, L., Bovolo, F., Benediktsson, J.A. (Eds.), Image and Signal Processing for Remote Sensing XXIII. SPIE, p. 3. https://doi.org/10.1117/12.2278218

Marston, Z.P.D., Cira, T.M., Hodgson, E.W., Knight, J.F., Macrae, I. V, Koch, R.L., 2020. Detection of stress induced by soybean aphid (Hemiptera: Aphididae) using multispectral imagery from unmanned aerial vehicles. J. Econ. Entomol. 113, 779–786. https://doi.org/10.1093/jee/toz306

Marston, Z.P.D., Cira, T.M., Knight, J.F., Mulla, D., Alves, T.M., Hodgson, E.W., Ribeiro, A. V, MacRae, I. V, Koch, R.L., 2022. Linear support vector machine classification of plant stress from soybean aphid (Hemiptera: Aphididae) using hyperspectral reflectance. J. Econ. Entomol. 115, 1557–1563. https://doi.org/10.1093/jee/toac077

Martins, G.D., Galo, M. de L.B.T., Vieira, B.S., 2017. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 5395–5403. https://doi.org/10.1109/JSTARS.2017.2737618

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. Biosyst. Eng. 114, 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009

Osco, L.P., Marques Ramos, A.P., Saito Moriya, É.A., Souza, M. de, Marcato Junior, J., Matsubara, E.T., Imai, N.N., Creste, J.E., 2019. Improvement of leaf nitrogen content inference in Valencia-orange trees applying spectral analysis algorithms in

UAV mounted-sensor images. Int. J. Appl. Earth Obs. Geoinf. 83, 101907. https://doi.org/10.1016/j.jag.2019.101907

Prabhakar, M., Gopinath, K.A., Kumar, N.R., Thirupathi, M., Sravan, U.S., Kumar, G.S., Siva, G.S., Meghalakshmi, G., Vennila, S., 2022. Detecting the invasive fall armyworm pest incidence in farm fields of southern India using Sentinel-2A satellite data. Geocarto. Int. 37, 3801–3816. https://doi.org/10.1080/10106049.2020.1869330

R Core Team, 2021. R: A language and environment for statistical computing.

Ragsdale, D.W., Landis, D.A., Brodeur, J., Heimpel, G.E., Desneux, N., 2011. Ecology and management of the soybean aphid in North America. Annu. Rev. Entomol. 56, 375–399. https://doi.org/10.1146/annurev-ento-120709-144755

Ragsdale, D.W., McCornack, B.P., Venette, R.C., Potter, B.D., Macrae, I. V., Hodgson, E.W., O’Neal, M.E., Johnson, K.D., O’Neil, R.J., Difonzo, C.D., Hunt, T.E., Glogoza, P.A., Cullen, E.M., 2007. Economic threshold for soybean aphid (Hemiptera: Aphididae). J. Econ. Entomol. 100, 1258–1267. https://doi.org/10.1093/jee/100.4.1258

Ramos, A.P.M., Gomes, F.D.G., Pinheiro, M.M.F., Furuya, D.E.G., Gonçalvez, W.N., Marcato Junior, J., Michereff, M.F.F., Blassioli-Moraes, M.C., Borges, M., Alaumann, R.A., Liesenberg, V., Jorge, L.A. de C., Osco, L.P., 2022. Detecting the attack of the fall armyworm (*Spodoptera frugiperda*) in cotton plants with machine learning and spectral measurements. Precis. Agric. 23, 470–491. https://doi.org/10.1007/s11119-021-09845-4

Reay-Jones, F.P.F., Greene, J.K., Toews, M.D., Reeves, R.B., 2009. Sampling stink bugs (Hemiptera: Pentatomidae) for population estimation and pest management in southeastern cotton production. J. Econ. Entomol. 102, 2360–2370. https://doi.org/10.1603/029.102.0643

Reisig, D., Godfrey, L., 2006. Remote sensing for detection of cotton aphid– (Homoptera: Aphididae) and spider mite– (Acari: Tetranychidae) infested cotton in the San Joaquin Valley. Environ. Entomol. 35, 1635–1646. https://doi.org/10.1093/ee/35.6.1635

Rhodes, M.W., Bennie, J.J., Spalding, A., ffrench-Constant, R.H., Maclean, I.M.D., 2022. Recent advances in the remote sensing of insects. Biol. Rev. 97, 343–360. https://doi.org/10.1111/brv.12802

Ribeiro, A.V., Cira, T.M., MacRae, I.V., Koch, R.L., 2022. Effects of feeding injury from *Popillia japonica* (Coleoptera: Scarabaeidae) on soybean spectral reflectance and yield. Front. Insect Sci. 2:1006092. https://doi.org/10.3389/finsc.2022.1006092

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring Vegetation Systems in the Great Plains with ERTS (Earth Resources Technology Satellite), in: Proceedings of 3rd Earth Resources Technology Satellite Symposium. pp. 309–317.

RStudio Team, 2021. RStudio: integrated development environment for R.

Santos, A. dos, Oumar, Z., Arnhold, A., da Silva, N., Oliveira Silva, C., Zanetti, R., 2017. Multispectral characterization, prediction and mapping of *Thaumastocoris peregrinus* (Hemiptera: Thamascoridae) attack in *Eucalyptus* plantations using remote sensing. J. Spat. Sci. 62, 127–137. https://doi.org/10.1080/14498596.2016.1220873

Vanegas, F., Bratanov, D., Powell, K., Weiss, J., Gonzalez, F., 2018. A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. Sensors (Switzerland) 18. https://doi.org/10.3390/s18010260

Yang, Z., Rao, M.N., Elliott, N.C., Kindler, S.D., Popham, T.W., 2009. Differentiating stress induced by greenbugs and Russian wheat aphids in wheat using remote sensing. Comput. Electron. Agric. 67, 64–70. https://doi.org/10.1016/j.compag.2009.03.003

Yuan, L., Bao, Z., Zhang, H., Zhang, Y., Liang, X., 2017. Habitat monitoring to evaluate crop disease and pest distributions based on multi-source satellite remote sensing imagery. Optik (Stuttg) 145, 66–73. https://doi.org/10.1016/j.ijleo.2017.06.071

Yuan, L., Zhang, J., Shi, Y., Nie, C., Wei, L., Wang, J., 2014. Damage mapping of powdery mildew in winter wheat with high-resolution satellite image. Remote Sens. (Basel) 6, 3611–3623. https://doi.org/10.3390/rs6053611

Zhang, Q., Li, B., Thau, D., Moore, R., 2015. Building a better urban picture: combining day and night remote sensing imagery. Remote Sens. (Basel) 7, 11887–11913. https://doi.org/10.3390/rs70911887