RESEARCH ARTICLE

Population Dynamics and Estimation of Damage of the Spittlebug *Aeneolamia varia* on Sugarcane in Colombia by Using remote Sensing and Machine Learning Tools

Joaquin Guillermo Ramirez-Gi[1] · William A. León-Rueda[1] · Mauricio Castro-Franco[2] · German Vargas[3]

Abstract The spittlebug (*Aeneolamia varia*) is one of the most important sugarcane pests in Colombia, where a recent increase in population and distribution specially in southwestern Colombia have led to the need for new technologies for integrated pest management. The objectives of this study were to determine the spatial distribution of this pest in commercial sugarcane fields and to validate machine learning (ML) tools for indirect injury detection and impact on yield (damage) using satellite images. This study was carried out in fields grown with the CC 01-1940 variety in El Cerrito, Valle del Cauca, Colombia, where systematic sampling of the populations (number of adults and nymphs per stem) was carried out. The spatial aggregation and distribution were determined using Moran's index and point patterns, sequence observations, and analysis with distance indicators (Sadie). The indirect injury detection and quantification of the impact on production were carried out with a ML approach using satellite image products with 10 m spatial and five days temporal resolutions, obtained from a Sentinel-2 sensor using Google Earth Engine. The results indicated that spittlebug populations had an aggregate spatial behavior and high spatial dependence. In addition, the ML algorithms predicted spittlebug injury, and the effect on production was estimated at 26.4 tons of cane per hectare, which represented a 17% reduction in the expected yield. The use of spatial analysis and remote sensing tools are an alternative for indirect detection of injury and for understanding population dynamics of the pest in sugarcane, so they can become instrumental for decision-making on an integrated pest management program.

Keywords Digital agriculture · Artificial intelligence · Prediction capacity · Integrated pest management · Sampling · Cercopidae

✉ Joaquin Guillermo Ramirez-Gi
jgramireg@unal.edu.co

[1] Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia Sede Bogotá, Bogotá, Colombia

[2] Precision Agriculture Research Group, FCAyRN, Universidad de los Llanos, Villavicencio, Colombia

[3] Department of Entomology and Nematology, University of Florida, Homestead, USA

Introduction

The global sugarcane (*Saccharum officinarum* L.) agroindustry has a large socioeconomic impact given its many products, including the production of sugar, "panela" (unrefined sugar), electricity, biofuels, paper, among others (Aguilar-Rivera 2019). In Colombia, sugar production is concentrated mainly in the southwest, in the departments of Valle del Cauca, Cauca, Quindío, Caldas, and Risaralda (Asocaña 2020). This sector represents 0.7% of the gross domestic product (GDP) *per capita* and generates more than 180,000 direct and indirect jobs, benefiting more than 1.2 million families in Colombia (Asocaña 2020).

Currently, a series of phytosanitary problems threatens the sugarcane agroindustry in Colombia, where the spittlebug *Aeneolamia varia* (F) (Hemiptera: Cercopidae) stands out (Vargas and Gutiérrez 2017). In southwestern Colombia, this pest is emerging, increasing its population and distribution, and presenting a considerable risk given its impact on production, which can be up to 25% of biomass production (Salazar and Proaño 1989), without mentioning the costs

associated with pest management and the environmental impact from the use of chemical insecticides (Vargas and Gutiérrez 2017). The injury of spittlebugs in sugarcane is associated with necrosis, portions of reddish brown spots and in advanced stages tissue death (Vargas and Gutiérrez 2017). This kind of injury it is related to nymphs and adults that can inject toxins into plant stems (Taliaferro et al. 1967).

Generally, the implementation of an integrated pest management program entails: (1) sampling and monitoring; (2) determination of the impact on the crop in terms of injury, damage, and loss; (3) application of treatments and evaluation of management practices; (4) forecasts of future population dynamics; and (5) implementation of site-specific agriculture (Brenner et al. 1998; Arbogast et al. 2000; Binns et al. 2000; Ferguson et al. 2003). Monitoring and subsequent spatial analysis are key to generating basic information on population dynamics, species detection, behavior, and pest distribution (Binns et al. 2000; Deleon et al. 2017; Arias et al. 2019).

Spatial analysis of sugarcane monitoring has optimized sampling schemes and, therefore, the integrated management of important arthropod pests, such as the weevil *Sphenophorus levis* Vaurie (Coleoptera: Curculionidae) in Brazil (Pavlu and Molin 2016); melolonthidae species (Coleoptera: Scarabaeidae) in the USA (Cherry 1984) and Australia (Allsopp and Bull 1989); and the sugarcane borer *Diatraea saccharalis* F (Lepidoptera: Crambidae) in the USA (Schexnayder Jr et al. 2001).

Remote sensing can optimize the spatio-temporal characterization of a pest attack, in a quickly, accurately, cheaply, and noninvasively ways (Khdery 2021). Specifically, it obtains key information on the interactions between pest attack, the structural changes of the crop, and its effect on the response in specific segments of the electromagnetic spectrum (Hatfield and Pinter Jr 1993). Currently, the combination of the Sentinel-2 A + B satellite constellation (European Space Agency—ESA—Copernicus Programme) and the massive processing platform Google Earth Engine (GEE) generates, validates, and updates numerous spectral monitoring tools for crops on multiple scales (Amani et al. 2020; Segarra et al. 2020). On the other hand, in the last decade numerous studies have reported on the usefulness of machine learning (ML) algorithms to data processing and analysis in the indirect detection of crop injury (Pardede et al. 2020).

The integration of ML tools and remote sensors to detect the effect of pests on sugarcane is incipient. However, recent work has focused on the detection of phytosanitary problems, stress conditions, and yield components (Abdel-Rahman and Ahmed 2008). In this regard, several studies have reported promising results in the prediction of plant growth and final yield (Natarajan et al. 2016; Ghazvinei et al. 2018), identification and mapping of weeds with spectral responses (Souza et al. 2020), and disease recognition with photographic images (Militante et al. 2019). Specifically, in analyses associated with spittlebugs in sugarcane, the use of a hybrid artificial intelligence model for the analysis of population parameters based on climate variables has been reported by Figueredo et al. (2021). In addition, several studies have reported that ML algorithms increase efficiency in time, precision and costs of injury detection, loss quantification, species identification, and monitoring (He et al. 2019; Partel et al. 2019; Rustia et al. 2020).

No studies are available yet, to the best of the authors knowledge, that validate the combination of ML algorithms and remote sensors in relation to spatial distribution and behavior of *A. varia* in sugarcane. Similarly, these technologies have not been used to classify the presence/absence of spittlebug injury and the subsequent damage, that is, its impact on sugarcane productivity. Finally, the quantification of the effect of spittlebug injury, in addition to generating key information on the affected productivity components, will generate a technical and economic base that will optimize integrated management. The objectives of this study were: (1) to determine and characterize the spatial–temporal dynamics of spittlebug populations and infestation in sugarcane crops; (2) to validate the use of ML algorithms for remote detection of spittlebug injury on sugarcane, using Sentinel 2 A + B satellite images; and (3) to quantify the effect of spittlebug damage on sugarcane yield.

Materials and Methods

Characterization of the Assessed Sugarcane Fields

This study was carried out in two 30 ha commercial fields (Field 1 and 2) planted with the sugarcane variety CC 01-1940 in a second ratoon (Fig. 1a). They were in the municipality of El Cerrito, Valle del Cauca, Colombia (latitude north: 3.693, longitude west: − 76.306; elevation: 960 m). The average monthly temperature varies between 19 and 30 °C, with minimums of 17 °C and maximums close to 32 °C, annual precipitation with bimodal distribution, and accumulated monthly values between 20 and 160 mm. The relative humidity is between 60 and 80%, and wind speeds are between 2 and 14 km h^{-1}. The agroecological zone is 6H1, which indicates less than 200 mm per year of excess precipitation and a predominantly sandy soil texture (Carbonell González 2011). All agronomic management practices were carried out according to the instructions of the technical staff in charge. The mechanical harvest was carried out for 13.1 months (06-01-2020), and the spatial obtaining yield was represented based on map (TCH: tons of cane per hectare) with a spatial resolution of 1 m, visualized using
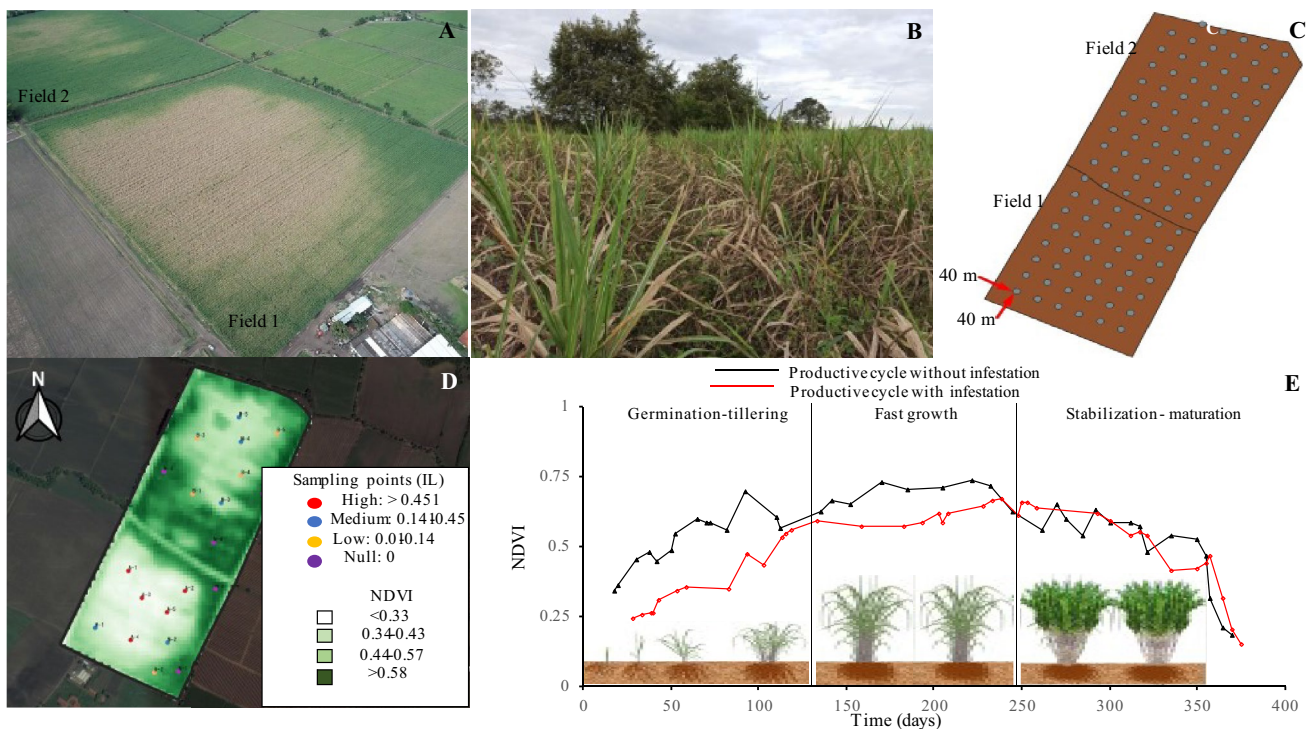
Fig. 1 Design and work scheme. **a** Fields 1 and 2 affected by *Aeneolamia varia* infestation. **b** Plant injury after *A. varia* attack, where chlorosis and necrosis of foliar tissue are evident. **c** Pest population sampling grid, 40 m×40 m. **d** Sampling points to measure crop development variables at 10 months of age. Infestation levels (IL) considered were: null (0 IL and >0.58 NDVI), low (0.01–0.14 IL and 0.44–0.57 NDVI), medium (0.141–0.45 IL and 0.34–0.43 NDVI), and high (>0.451 IL and <0.33 NDVI). IL was determined as the mean of adults and nymphs per stem. **e** NDVI index through a former crop cycle without pest infestation and a subsequent one with pest infestation. NDVI data were acquired from historical remote imagine from Sentinel-2 using Google Earth Engine. Normalized difference vegetation index (NDVI)

the QGIS version 3.16 geographic information system tool (QGIS Development 2016).

Monitoring Spittlebug Populations

At the middle of 2019, a spittlebug attack was detected in the fields described above (Fig. 1a and b), with a crop of 5 months of age. So, monitoring of pest's immatures (nymphs) and adults was carried out. Then, a systematic sampling was designed in a regular georeferenced 40×40 m grid, selecting 51 and 86 sampling points for fields 1 and 2, respectively (Fig. 1c). Field monitoring was carried out on May 25 and 31 and July 2, of 2019.

Variables associated with the population level of adults and nymphs were evaluated at each sampling point during three monitoring sessions. Each point was represented by row meter transects (10 m), where the number of stems and the number of adults present in the foliage were determined; then, the root zone was checked for spittle masses, assuming each one contained a nymph. Finally, estimates of the infestation were made by calculating the number of nymphs/stem and adults/stem (Vargas and Gutiérrez 2017). Additionally,

injury to the foliar tissue on the plants was evaluated, quantifying it as a dichotomous variable (presence–absence). The injury at the foliar level was characterized as portions of reddish brown spots, chlorosis, and necrosis in a more advanced stages tissue death (Fig. 1b) (Vargas and Gutiérrez 2017).

Spittlebug Population Inference and Level of Injury

With samplings executed in three different times (described above), a data set was obtained and used for estimates of population parameters using descriptive statistic methods. For pest population variables (number of nymphs and adults) and infestation (nymphs/stem and adults/stem), the mean, mode, minimum, maximum, standard deviation, variance, and coefficient of variation (CV) were estimated. In addition, with the aim to evaluate whether there were population changes over time, an one-way analysis of variance and subsequent means comparisons were performed using Tukey's test ($P < 0.05$). The Levene and Kolmogorov–Smirnov criteria were used to evaluate the homoscedasticity and normality of the data, respectively. Analyses were run in the free statistical software R version 4.2.1 (R Development Core Team 2022).

Analysis of the Spatio-temporal Dynamics of the Spittlebug Populations

The analysis of the spatial–temporal dynamics of nymphs and adults was carried out in two steps. The first step determined the aggregation (random, aggregate, or uniform) and spatial dependence of spittlebug populations using the Moran index (MI) (Moran 1948). In the second step, the spatial distribution was modeled, and since the number of nymphs and adults were counting types (discrete) and distributed randomly, spatial tools of the point and area pattern type were used (Cressie 1991).

The methods used in this second step were: (1) analysis distribution patterns using the kernel density estimator, where the points were a function of the coordinates (xy), and the mark was a function of the value of the population count (nymphs and adults). This process was implemented with the statistical software R and the functions of the spatstat package (Baddeley et al. 2015). (2) Area patterns using count sequence observations (population of nymphs and adults) in space (coordinates) based on the monitoring carried out at each sampling point in the field, implemented with R and the functions of the epiphy package (Gigot 2018). (3) Spatial analysis with distance indicators (Sadie) (Perry 1995; Li et al. 2012), implemented with R and the functions of the epiphy package (Gigot 2018). The Sadie index is designed for individual count analysis, obtaining an aggregation index (Ia) and associating the presence of neighbors with a high presence (Vi) or absences (Vj) (Perry 1995; Li et al. 2012). Likewise, the colors in the points provide a visual representation that is related to clusters associated with the size and dimension of the studied phenomenon (Perry et al. 1999a, b).

Acquisition of Multi-band Satellite Images with High Temporal and Satellite Resolution and Determination of Vegetation Indices

Our approach was based on evaluating two sugarcane production cycles: (1) without infestation and (2) with spittlebug infestation. The average length/cycle was 13 months. For both, monitoring was carried out using satellite images from planting to harvest. The time window evaluated was from years 2017 to 2020, specifically the first crop cycle occurred between October 2017 and November 2018, while the cycle of high infestation, between December 2018 and January 2020.

Multispectral satellite images were the source of the overall set of predictor variables. These images were obtained from the Sentinel-2 A + B constellation system supported by the "Copernicus Land Monitoring studies" mission (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a), with a spatial and

temporal resolution of 10 m and 5 days, respectively, for October 2017 to January 2020. The selected bands were: 2-blue (0.45–0.52 µm), 3-green (0.54–0.57 µm), 4-red (0.65–0.68 µm), 8-NIR1 (0.78–0.90 µm), and 11-SWIR 1 (1.56–1.65). The images were acquired from the planetary-scale platform for Earth science data and analysis—Google Earth Engine (GEE) (Amani et al. 2020), using own JavaScript code.

The multi-step scheme that was executed on the GEE platform consisted of: (1) selection of the luck area and obtaining all the available images for the 2017–2020 period; (2) cloud mask and threshold to select products that meet quality and information parameters; (3) clipping, band extraction and calculations of pixel values associated with reflectance to determinate the vegetation indices described in the next section; (4) visualization of vegetation indices for the lots and extraction of characteristics for the sampling points; and (5) export information. Additionally, since band 11—SWIR 1 had a spatial resolution of 20 m, an interpolation and resampling process was carried out at 10 m, supported by the 10 m resolution bands. For this process, a convolutional neural network (Wu et al. 2022) was optimized and applied, implemented in the free Python programming language using the TensorFlow and Keras libraries (Brownlee 2016).

In total, 39 (for 2017–2018) and 31 (For 2018–2020) Sentinel-2 satellite images were obtained for the fields, divided into two production cycles: (1) without spittlebug affectation (2017–2018) and (2) with spittlebug affectation (2018–2020), respectively. With the images, previously reported vegetation indices with a high potential for detecting alterations in plants caused by arthropod pests were calculated (Table 1). Additionally, blue, green, red, and near-infrared (NIR) bands were obtained. Subsequently, these products were projected to the evaluated fields. This process was executed with R with functions in the LSRS package (Sarparast 2018).

Indirect Detection of spittlebug Injury Using Machine Learning Tools

To indirectly detect the spittlebug injury in the host, a set of multi-band satellite images (described above) were used that had a temporal correspondence after field pest population evaluations between the fifth day and 15 days after the last field evaluation (May 25 and 31 and July 2, 2019). Then, the indices described in Table 1 were determined, and the bands were extracted for use as predictor variables. Healthy plants (absence of injury), soil, and injured plants were used as classes to be predictors. The presence class was obtained by extracting the mean of the sampled plants at any of the sampling points (Supplementary Information 1 A and B), indicating presence class as plants presenting some level

Table 1 Mathematical formulas of several vegetation indices used to plant indirect pest injury detection

Name	Equation	References
Normalized difference vegetation index (NDVI)	$NDVI = {}^{R}_{NIR} - {}^{R}_{RED}/R_{NIR} + {}^{R}_{RED}$	Genc et al. (2008)
Green normalized difference vegetation index (GNDVI)	$GNDVI = {}^{R}_{GREEN} - {}^{R}_{RED}/{}^{R}_{GREEN} + {}^{R}_{RED}$	Sudbrink et al. (2015)
Two-band Enhanced Vegetation Index (EVI2)	$EVI2 = 2.5^{*}(({}^{R}_{NIR} - {}^{R}_{RED})/(R_{NIR} + 2.4^{*R}_{RED^-} + 1))$	Mondal (2011)
Ashburn vegetation index (AVI)	$AVI = {}^{R}_{NIR} - {}^{R}_{RED}$	Sudbrink et al. (2015)
Soil-adjusted vegetation index (SAVI)	$SAVI = ({}^{R}_{NIR} - {}^{R}_{RED})(1 + L)/{}^{R}_{NIR} + {}^{R}_{RED} + {}^{L1}$	Yang et al. (2007)
Normalized difference water Index (NDWI)	$NDWI = {}^{R}_{NIR} - {}^{R}_{SWIR}/{}^{R}_{NIR} + {}^{R}_{SWIR}$	Huo et al. (2021)
Atmospheric resistant vegetation (ARVI)	$ARVI = ({}^{R}_{NIR} - (2^{R}_{RED-BLUE}))/({}^{R}_{NIR} + (2^{R}_{RED-BLUE}))$	Yang et al. (2005)

Near infrared (NIR)(0.85–0.88 nm); red (0.64–0.67 nm), green (0.53–0.59 nm); blue (0.45–0.51 nm); short-wave infrared (SWIR1)(1.75–1.65 nm) = Landsat bans. R: surface reflectance values. L: to adjust the background = 1

of injury (as described above), while absence means plants with no evidence of injury.

For the classification process of the previous classes, the ML approach was used. Specifically we selected the random forest (RF) classifier for simplicity, low computational cost and effectiveness (Pal 2005). We implemented an assembly model of the bootstrap Bagging aggregation (Breiman 1996). The data were randomly divided into two data sets: (1) training (75%) and (2) testing (25%), and in each run of the model, a seed was placed to stabilize the predictions and achieve reproducibility in the results.

Different variations of input data and parameters were made within the RF algorithm, seeking a balance between robustness, significance, and computational performance. A set of variable combinations (bands and bands + indices) and classes was evaluated. At the algorithm level, variations were made in the number of trees, number and depth of nodes (1–4000), and the hyperparameter alpha (0–20) (Henao-Rojas et al. 2021). Additionally, the evaluated classes were balanced, incorporating artificial class weights in the RF classifier algorithm (Liaw and Wiener 2002).

The classification result was evaluated using five statistical parameters: (1) the receiver operation characteristic (ROC), comparing the values on the scalar value of the area under the curve (AUC), a measure of the quality of the classification (Cortes and Mohri 2003); (2) confusion matrix (Bekkar et al. 2013; Mueller and Guido 2016), where the rows corresponded to the true classes and the columns to the predicted classes. With each matrix, the global accuracy (GA) was calculated using the ratio between the total number of correct samples over the total number of samples (Hasmadi et al. 2009); (3) the Kappa index; (4) inference error using a stratified 10 K-fold cross-validation (1-inference error = precision); and (5) significance ($P < 0.05$).

The supervised classification and the evaluation of the results were developed by building code in the free software R using functions from the caret (Kuhn et al. 2020); random forest (Liaw and Wiener 2018); Car (Fox et al. 2020); pROC

(Robin et al. 2021); randomForestExplainer (Paluszynska and Biecek 2019); and Raster (Hijman 2016) packages.

Estimation of Spittlebug Infestation Effects on Yield Using Indirect Detection of Injury and a Machine Learning Approach

With the objective of quantifying some productivity parameters prior to harvest, a stratified sampling was designed on the 10th month of crop development (20 of September 2019) (Fig. 1d). The infestation level (IL) was associated with the normalized differential vegetation index (NDVI) according to the distribution of quartiles as follows: (1) high infestation and low NDVI values (Q4: > 0.451 IL and < 0.33 NDVI); (2) mean infestation and intermediate NDVI values (Q2–Q3: 0.141–0.45 IL and 0.34–0.43 NDVI); (3) low infestation and medium high NDVI values (Q1–Q2: 0.01–0.14 IL and 0.44–0.57 NDVI); and (4) no infestation and high NDVI values (QI: 0 IL and > 0.58 NDVI) (for further details, see Fig. 1f). The IL for each evaluation point was obtained as the mean of the pixel value after joining the interpolation rasters, using the kernel density method (described above) for the variables nymphs/stem and adults/stem and functions of the raster library, implemented in the free software R (Hijman 2016).

The free application Avenza Maps (Avenza Systems Inc) was used for the geolocation of each sampling point. Then, the number, height and diameter of the stems were determined in transects of 10 m each. The data were subject to an analysis of variance and post hoc tests, using the Tukey test ($P < 0.05$), after verifying the normality and homoscedasticity of the data using the Levene and Kolmogorov–Smirnov criteria, respectively, using statistical free software R.

The second analysis aimed to evaluate the use of multivariate regression methods and an ML approach for predicting infestation effects on sugarcane yield using predictive variables obtained from satellite images described above. Our approach was based on the sensitivity of NDVI obtained at a resolution of 10 m from Sentinel-2 sensor images for the

characterization of phenological phases and productivity of sugarcane (Wang et al. 2019) (Fig. 1e).

With the 39 and 31 images obtained for the above-mentioned periods, the mean spectral responses of each index and band were calculated (Table 1) for each phenological phase for the two evaluated cycles (Fig. 1g) as predictor variables. The prediction variable was the yield, expressed in tons of cane per hectare (TCH) for the 2017–2018 (without infestation) and 2018–2020 harvests (with spittlebug infestation), obtained from the mechanized and visualized harvest with the raster format of continuous values using the free software QGIS.

The extraction of the values from the raster originated with the indices, and the specific bands [2-blue (0.45–0.52 µm), 3-green (0.54–0.57 µm), 4-red (0.65–0.68 µm), 8-NIR1 (0.78–0.90 µm), and 11-SWIR 1 (1.56–1.65]) were done using the raster library (Hijman 2016), while the yield values were obtained from the mechanic harvest map described before using aspatial join of the rgeos library, as the information was in representative polygons at each pixel (Bivand et al. 2020).

With this information, some parameters of spatial descriptive statistics were determined. The behavior of the spectral indices and bands at different points of crop development and their relationship with the TCH were estimated by calculating the centroids associated with each polygon of the production area, and the subsequent extraction of values, whose visualization was carried out using histograms, was done using the sf library of the free software R (Pebesma 2018). In addition, histograms were generated per infestation level to compare the behavior between sampling days and the spectral response in the crop.

An estimation model was fitted from the values of each pixel for the variable to be predicted (TCH) and the predictors (bands and indices obtained from spectral images), comparing two approaches. In the first approach, we used a generalized additive model (GAM) (Eq. 1). Initially, a selection process of non-autocorrelated variables was carried out using the Pearson or Spearman coefficient, eliminating those with values less than 0.8. Subsequently, the selection of predictive variables was carried out with a stepped selection method, using the Akaike information criterion (AIC) until the variables were correlated with the residuals (Efron et al. 2004). The procedures were developed using R and the functions of the gam library (Hastie 2020).

$$\text{TCH} = \beta_0 + f_1(xs_1) + f_2(xs_1) + \cdots + f_m(xs_m), \quad (1)$$

where TCH is the tons of cane per hectare; β_0 is the intercept; and $f(xs)$ is the standardized vegetation bands and indices according to z transformation methods using the expression reported in Eq. 2.

$$Zi = \frac{\chi i - \mu}{\sigma X}, \quad (2)$$

where $\chi i =$ the raw variable value, $\mu =$ mean, and $\sigma =$ standard deviation.

In the second approach, we implemented ML algorithms. We selected a regularized linear regression using the methods Lasso (stands for Least Absolute Shrinkage and Selection Operator) regression model because this method selects variables, excluding less relevant predictors (Friedman et al. 2010). The penalty of this model was controlled by evaluating a set of values of the hyperparameter λ (0–20), selecting the λ value that gave rise to the best model through cross-validation using the statistical parameter of the mean square error (MSE). The calibration and validation of the model were developed with a data set, 75 and 25%, respectively. The procedures were developed using the software R and the functions in the glmnet library (Friedman et al. 2021).

The evaluation of the two models was carried out using the coefficient of determination (adjusted R^2), AIC, and root mean square error (RMSE). The procedures were developed using the free software R. The results of the selection of variables with the Lasso regression method (described above) showed that the biomass production (TCH) was highly influenced by the level of spittlebug infestation (adults and nymphs per stem) and the indirect injury detection associated with NDVI values. So, the TCH and infestation were fitted to a second-order linear and quadratic model as a function of the infestation level and the NDVI, respectively.

The TCH and infestation data for each sampling point were obtained by extracting the values according to the coordinates of the harvest map and the mean of the three samplings. The fit of second-order linear and quadratic models was evaluated with statistical significance ($P < 0.05$), coefficient of correlation (r), determination (R^2), and mean square error (MSE). Likewise, for all models, the assumptions of normality, independence, constant variances, and linearity were guaranteed. The procedures were developed using the statistical free software R.

Results

Spittlebug Infestation Levels and Space–Time Population Dynamics

The monitoring of adult spittlebug populations and their level of infestation showed high variation over time, represented by values of the CV between 102.5 and 234%. The highest average populations were present in the first and second sampling times, with a reduction in the third evaluation. In this sense, maximum population values of 110 adults were observed in a linear row meter, along with maximum

infestation values of 73 adults per stem, for all samples (Table 2). This variation gave rise to significant differences depending on the sampling time (Fig. 2).

For the nymph population and infestation, the means were similar in all three samplings but with high variation, as represented by the CV and high number of atypical values, higher than Q4 of the box-and-whisker plot (Table 2, Fig. 2a and c). In some sampling points, maximums of 151 nymphs were found, with infestation levels of 78 nymphs per stem. This high variation resulted in no statistical differences, depending on the sampling time (Table 1 and Fig. 2b and d).

For the spatial dynamics, it was found that, according to the Moran index, the adult populations had an aggregate distribution, highly aggregated distribution, and medium aggregation, with a random tendency, in samples 1, 2, and 3, respectively (Table 1). On the other hand, the populations of nymphs presented a very high, medium, and medium–low aggregation for the three samplings, respectively (Table 1) (Fig. 1a).

For the spatial distribution, populations of nymphs and adults were generally grouped in areas within the lot with variations depending on the sampling (Figs. 1a and 3). The higher concentrations were in the central and southwestern areas of the fields, with a tendency toward spatial aggregation. The three methods of spatial analysis used found coincidence in the above-mentioned. On the other hand, the interpolation with the kernel method assigned higher values of density or grouping, depending on the state type of the insect and the sampling time (Fig. 3a). Likewise, the area pattern method grouped centroids with population values for adults and nymphs between 1 and 7 individuals (Fig. 3b). The Sadie index was able to differentiate areas with a high risk of finding populations of adults and nymphs (red, Fig. 3c) with an aggregate distribution, those areas with

low levels (blue dots, Fig. 3c), and the areas with movement of insects toward new areas (white color, Fig. 3c).

For the nymphs, higher levels of population grouping were observed in the first and third sampling, with greater intensity in the southeast, center, and northeast (Fig. 3), as corroborated by the interpolation using the kernel density (Fig. 3a). This result was corroborated with the Sadie index, which showed that, in these areas, there was aggregation and a greater probability of finding nymphs (Fig. 3c). The densities of adults were considerably higher in the first sampling, with a tendency to concentrate toward the edge of the lot in the southwest but with a more random distribution (Fig. 3b). Later, in samplings 2 and 3, although the population level decreased, the aggregation increased, moving from the edges to the central areas in the fields (Fig. 3).

Remote Spittlebug Injury Detection Using Satellite Imagery and ML Tools

It was found that the ability to correctly discriminate the classes associated with injured plants associated with spittlebugs, without injury, and soil using vegetation indices and bands obtained from satellite images varied depending on the number of variables used as predictors, the sampling time, and the optimization of the RF algorithm according to the number of nodes, minimum depth, and hyperparameter λ (Fig. 4 and Table 3).

For the RF algorithm, the best prediction was based on a balance between the accuracy, the values of the kappa statistic, the ROC-AUC and the significance for the training and testing data set and was obtained with a combination of predictor variables between 4 and 10, a number of nodes between 120 and 334, a minimum depth of nodes between

Table 2 Descriptive statistics and population of nymphs, adults per meter of crop row and infestation expressed as nymphs and adults per stem in sugarcane commercial plots under natural infestation of *Aeneolamia varia*

Sampling	Variable	Average	Mode	Median	Minimum	Maximum	CV (%)	Moran index[a]
1	Number of nymphs	12.4	0.0	3.0	0.0	77.0	146.1	0.8
	number of adults	22.4	0.0	18.0	0.0	110.0	102.5	0.8
	Nymphs/stem	5.1	0.0	1.0	0.0	28.0	142.5	na
	Adults/stem	9.5	0.0	7.0	0.0	73.0	121.3	na
2	Number of nymphs	14.0	0.0	6.0	0.0	115.0	150.0	0.5
	number of adults	2.0	0.0	0.0	0.0	34.0	224.2	0.8
	Nymphs/stem	6.8	0.0	3.0	0.0	56.0	152.0	na
	Adults/stem	0.93	0.0	0.0	0.0	16.0	234.0	na
3	Number of nymphs	12.7	0.0	4.0	0.0	151.0	175.0	0.6
	number of adults	1.94	0.0	0.0	0.0	22.0	212.5	0.6
	Nymphs/stem	7.47	0.0	2.0	0.0	78.0	173.7	ma
	Adults/stem	1.1	0.0	0.0	0.0	12.0	209.2	na

CV coefficient of variation

[a]Values close to 0, less than −1, and close to 1 indicate a random, uniform, and aggregated spatial pattern, respectively (Moran 1948)
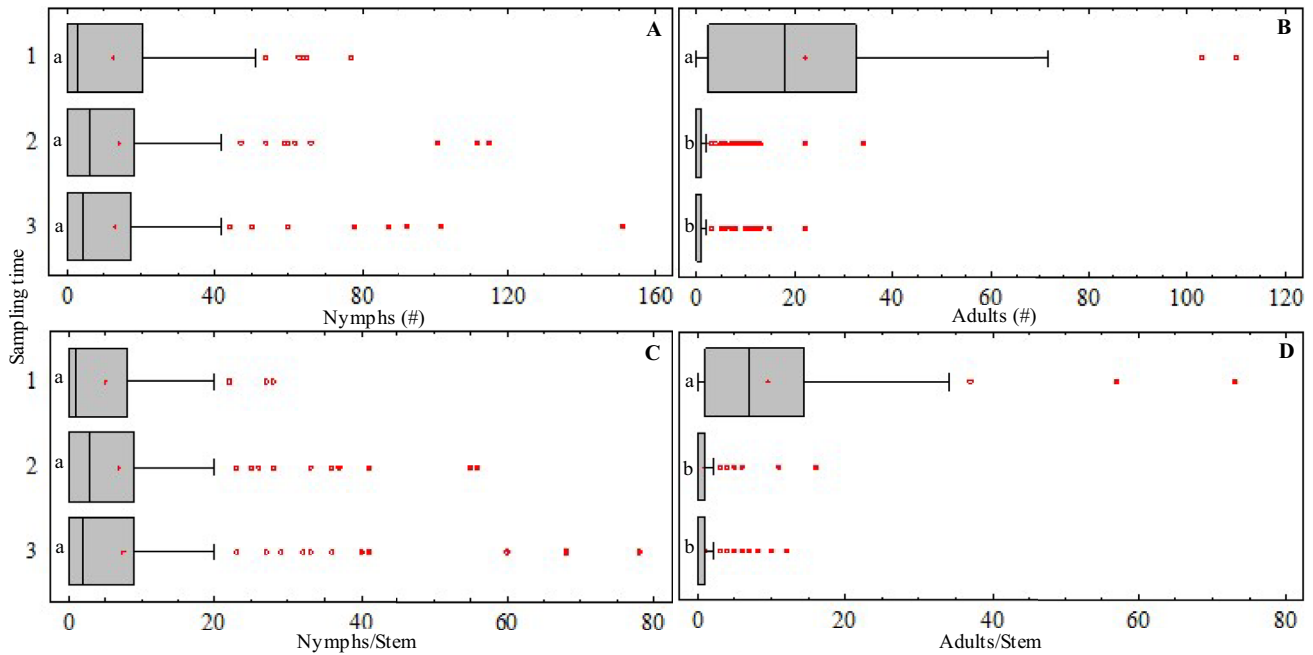
Fig. 2 Boxplot representation and significance of population level and infestation of immature and adult of *Aeneolamia varia* in sugarcane at different monitoring times (**1, 2, and 3**). No overlapping of the error bars of equal letters indicates significant differences (Tukey, $\alpha = 0.05$). **a** Nymph population per meter of crop row, **b** adult population per meter of crop row, **c** nymph's infestation expressed as nymphs/stem and **d** adult's infestation expressed as adults/stem

1.91 and 418, and hyperparameter λ values between 2 and 5 (Table 3).

For the discrimination of classes associated with injury associated with spittlebug adults or nymphs, it was found that the informative variables differed for each sampling time. First, a highly predictive capacity for the presence of spittlebugs injury in the NIR band was observed, followed by the AVI, EVI, and SAVI vegetation indices. Likewise, a good prediction for the absence of spittlebugs injury was obtained in the red, green, and blue bands. In the second sampling time, a low predictive capacity was observed for the bands and indices for pest injury, while for the absence of injury the more important predictors were the NDVI, AVI, ARVI, and NDWI vegetation indices. In the third sampling, only the green band had predictive importance for the presence of injury of spittlebugs, while for absence, the more important predictors were the EVI, NDVI, and SAVI vegetation indices and the blue and NIR bands. In the three samplings, the red, green, and blue bands presented predictive capacity for the soil class (Fig. 4a).

According to the values found in the confusion matrix (Fig. 4b), the results indicated that the trained RF model presented variations in the correct selection of classes as a function of time. In the first sampling, the presence of pest injury was moderately estimated, but the absence was underestimated since a higher value indicated a presence. On the other hand, in the second and third sampling times, the opposite occurred, and the classes associated with pest injury were underestimated.

The prediction accuracy of the evaluated classes, as well as the other evaluated statistics, varied depending on the number of classes, the combination of the predictive variables and the sampling time (Table 4). In general, a better fit was found using all bands and indices, all classes, and sampling times (Table 3). Based on the above, a best-case accuracy of 68, 70, and 72% was achieved for monitoring 1, 2, and 3, respectively (Table 3). Figure 4c represents the spatial distribution of the classification using the RF algorithm of the evacuated classes in the three samples. Here it is evident a considerable difference in the magnitude of the presence of injury associated with spittlebugs classified between the first sampling in comparison to the other two. Despite these differences, the prediction of presence of injury tended to be concentrated in similar areas during the three surveys and was highly associated with the population analyses of nymphs and adults using spatial statistic tools (Fig. 4). The soil class was identified with higher density in the second sampling, possibly because of the high infestation resulting from high population levels in those same areas during the first sampling (Table 2 and Fig. 4).

There was a correspondence between the density of pixels for each of the seven vegetation indices and the bands evaluated with respect to the level of infestation of nymphs and adults (Supplementary Information 1), explaining how the

Fig. 3 Analysis of the spatial distribution of population level of nymphs and adult spittle-bugs in sugarcane at different monitoring times. SADIE index range: Values of the index range > 1.5 indicate aggregation of the populations, values of the index range < 1.5 indicate non-aggregation and values between − 1.5 and 1.5 indicate areas of interception of the two phenomena. The size and color of the circles (red and blue) indicate the size, intensity, and dimension of the phenomenon (aggregated and non-aggregated populations) (Perry et al. 1999a, b)
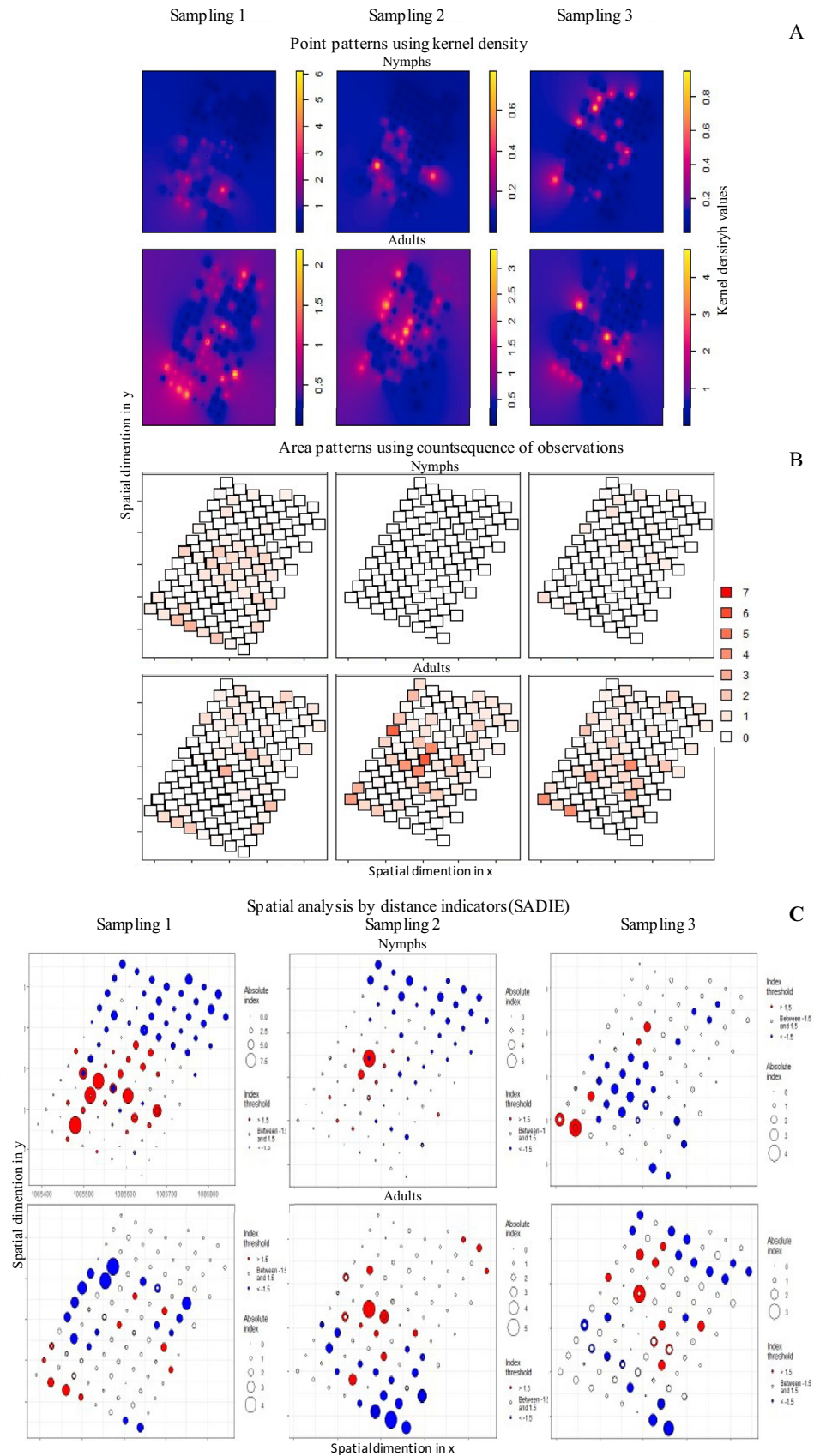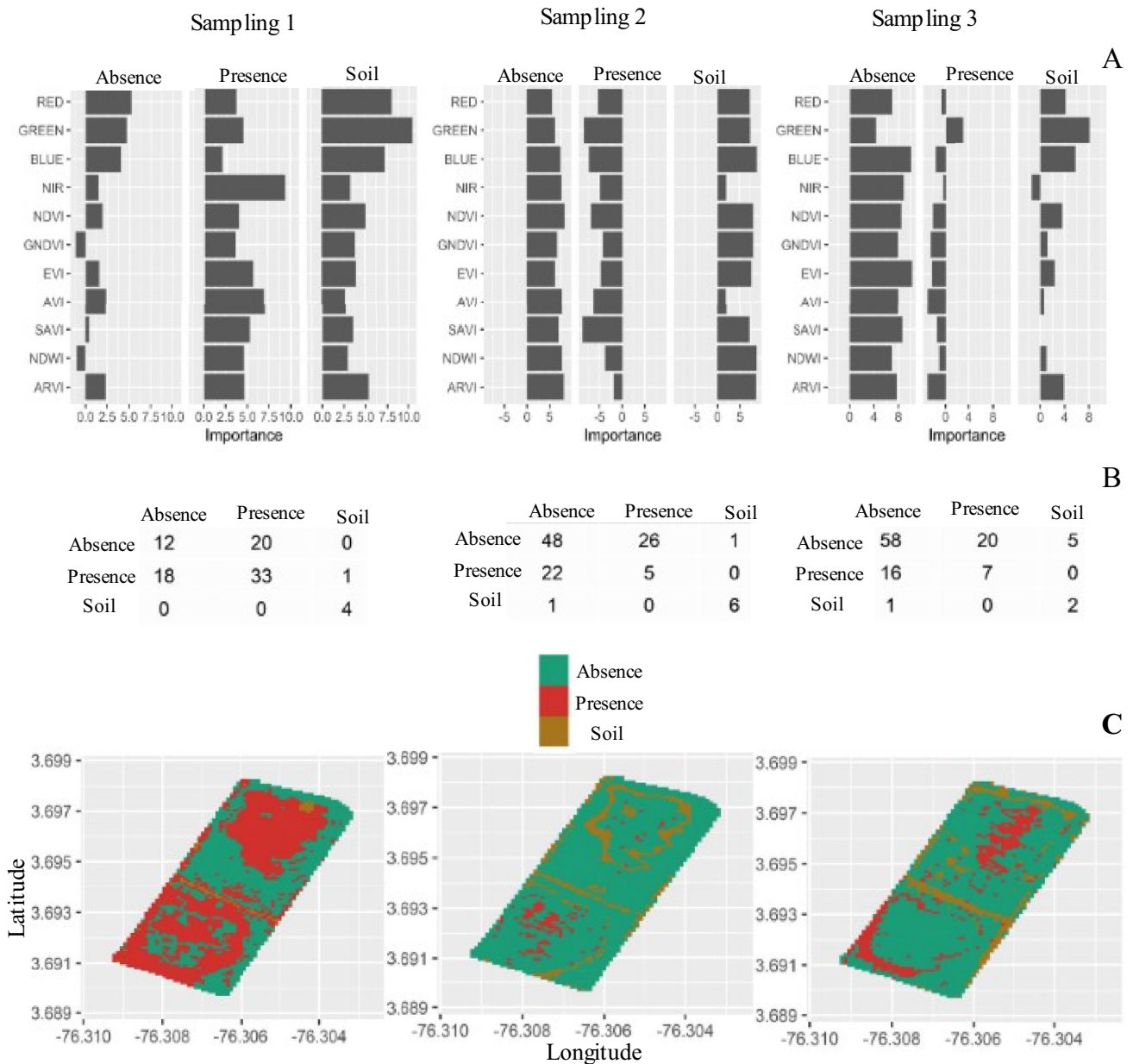
Fig. 4 Results of the classification of sugarcane plants with spittle-bugs injury using machine learning tools and spectral images as predictors. **a** Importance of variables on the discrimination of classes. **b** Confusion matrix and classification infestation of immature, rows correspond to the true classes and the columns to the predicted classes. **c** Adult and nymphs spittlebugs injury classification in sugar-cane at different monitoring times using the random forest algorithm. Normalized difference vegetation index (NDVI); Green normalized difference vegetation index (GNDVI); two-band Enhanced Vegetation Index (EVI2); Ashburn vegetation index (AVI); Soil-adjusted vegetation index (SAVI); Normalized Difference Water Index (NDWI); Atmospheric resistant vegetation (ARVI)

lower values in each index and band were associated with higher infestation levels, along with intermediate and high values with low and medium infestation levels (Supplementary Information 1).

Determination of Spittlebug Infestation and Its Effect on Yield

The more informative vegetation indices for discriminating differential values of TCH values were NDVI, GNDVI, EVI, AVI, and ARVI (Fig. 5a–c), explaining how the NDVI was highly consistent and sensitive in differentiating the different phenological phases of sugarcane during its crop cycle and

Table 3 Optimized parameters and statistics metrics of random forest machine learning method to ability for discrimination of classes and pest spittle detection in sugarcane

Sampling time	Model	Optimized parameters in the random forest algorithm				Statistical metrics in training/testing data set							
		mtry[1]	MNN	MND	Hα	Accuracy[2]		Kappa		Multi-class AUC[3]		p value[4]	
						Training[a]	Testing[b]	Training[a]	Testing[b]	Training[a]	Testing[b]	Training[a]	Testing[b]
1	Bands – all class	4	120	2.1	2	0.6	0.6	0.7	0.8	0.9	0.9	**a	**b
	Bands + Index – all class	10	221	3.1	3	0.7	0.7	0.8	0.8	0.9	0.9	***	***
	Bands + Index – selected class	9	234.1	4.2	4	0.7	0.6	0.7	0.7	0.9	0.9	**	**
2	Bands – all class	4	100	1.9	3	0.6	0.6	0.6	0.6	0.9	0.8	**	**
	Bands + Index – all class	10	220.8	3.2	3	0.7	0.7	0.7	0.7	0.9	0.9	***	***
	Bands + Index – selected class	8	200	2.8	5	0.7	0.7	0.7	0.7	0.9	0.9	***	***
3	Bands – all class	4	120	1.9	3	0.6	0.6	0.6	0.6	0.8	0.8	**	**
	Bands + Index – all class	11	320	2.3	5	0.6	0.7	0.6	0.7	0.9	0.9	***	***
	Bands + Index – selected class	9	240.1	3.3	4	0.8	0.9	0.6	0.9	0.9	0.9	***	***

[1]Number of variables available for splitting at each tree node (square root of the number of predictor variables). Mean number of nodes (MNN). Mean minimum depth (MND). hyperparameter alpha (Hα)

[2]Inference-error using stratified 10 K-fold cross-validation (1-inference error = accuracy.

[3]Determined parameter using area under receiver operating characteristic (ROC) curve (AUC).

[4]The asterisks in the last column indicate the significance ranking of the P values: *** = 0.001; ** + 0.01; * = 0.05

[a]Training data set

[b]Testing data set

Table 4 Ability of traditional and machine learning methods to predict the yield on sugar cane without and with infestation of *Aeneolamia varia*

Productive cycle	Method	RMSEP	AIC	R^2 (%)	Yield (TCH)				Computational times	
					Real	Predicted	Reduction by infestation		Algorithm	Time needed
							Real[1]	Predicted		
Without infestation	GAM	96.2	7,092	84.5	153.5	164.3	na	na	100×RDCV	2 min 10 s
	Lasso	28.9	3,153	90.7	153.5	159.2	na	na	Single CV	1 min
With infestation	GAM	112.1	8,812	82.1	118.8	126.4	24.7	17.1	100×RDCV	3 min 55 s
	Lasso	67.4	4,529	88.3	118.8	122.1	24.7	26.4	Single CV	1 min

GAM generalized additive model, *Lasso* least absolute shrinkage and selection operator, *RMSEP* root mean square error of prediction, *AIC* Akaike information criterion, R^2 coefficient of determination

[1]The reduction by infestation was determined by means of the following expression:

RI = (TCHn − 1 − (RR)) − TCHn, where

RI: reduction by infestation, TCH$_{n-1:}$ tons of sugarcane per hectare of the previous period without infestation; *RR* ratoon reduction, for the case of an initial cut ($n=1$) to a later one ($n=2$), it was assumed to be 10 TCH; y TCH$_n$: TCH of the period affected by the infestation

achieving a clear differentiation between the cycle without infestation and where the *A. varia* outbreak occurred in this commercial sugarcane fields (Fig. 1g). For the relationship of the mean values of the vegetation indices during the crop cycle with the TCH, lower levels coincided with the areas where there was less yield (Fig. 5a, b). These results coincided with the fact that higher numbers of pixels with high index values, such as NDVI, GNDVI, EVI, AVI, and ARVI, were found in the classes with high production (120–124.9 and 130–142 TCH) (Fig. 5a–c). The latter results show consistency of areas with a lower population level of nymphs and adults (Fig. 3) and areas classified by the RF algorithm as non-infested areas (Fig. 4).

At the 10th month of crop development the comparison of relationships between the productivity components related to the number, weight and diameter of stems, and based on the categories of infestation levels from the NDVI indices: none (0 IL and > 0.58 NDVI), low (0.01–0.14 IL and 0.44–0.57 NDVI), medium (0.141–0.45 IL and 0.34–0.43 NDVI), and high (> 0.451 IL and < 0.33 NDVI), presented statistical differences (Figs. 1c and 6a–c). High infestation levels led to a lower number, height, and diameter of stems. Between the medium and low categories, there were no considerable differences in the productivity components in terms of relative magnitudes; the losses from the effect of high levels of spittlebug infestation occurred in descending order: number of stems (~ 70% lower), height of stems (~ 48% smaller), and stem diameter (~ 33% smaller) (Fig. 6a–c).

For the prediction of production using the two approaches in this study, the traditional method using a GAM-type model was inferior to the Lasso regression (Table 4). The prediction of TCH for both methods was slightly overestimated, but the value did not exceed 10% of the real performance, with a better fit for the Lasso regression with respect to the GAM (Table 4).

For both crops cycles (with and without infestation), the methods were sensitive in detecting changes in the estimated yield based on the mean values of the spectral indices and bands obtained from satellite images during a cycle of approximately 13 months. The ML method was able to predict that the losses associated with the spittlebugs injury were 26.4 TCH, when high infestations of nymphs and adults per stem were compared with similar areas but without infestation, while the real value of the reduction from the affectation was 24.7 TCH, corresponding to a 17% reduction in the expected production, and where the estimated loss value with the GAM method was lower (17.1 TCH) (Table 4).

The simple linear models presented a good fit, indicating that the variation explained by the predictor variables was greater than a random effect (Fig. 7a–d). Negative relationships were found between infestation and plant productivity variables. In the case of the number of adults per stem, there was a first-order relationship between the infestation and TCH ($F = 145.23$; $gl = 1$, 135; $P < 0.001$; $r = -0.71$; $R^2 = 51.82$; MSE = 5.60; Fig. 7a), while, between the infestation and the NDVI index, the relationship was quadratic ($F = 141.8$; $df = 2$, $134 = P < 0.001$; $R^2 = 67.92$; MAE = 0.06; Fig. 7c). According to the relationship between adults per stem and TCH, an increase of one unit in the infestation would be associated with a decrease in TCH of 19.2%. In reference to the nymphs per stem, there was a first-order relationship between infestation and TCH ($F = 36.95$; df = 1, 135; $P < 0.001$; $r = -0.56$; $R^2 = 41.48$; MAE = 7.32; Fig. 7b), while between the infestation and the NDVI index the relationship was of the second order ($F = 25.95$; $df = 2$, 134 $P < 0.001$; $R^2 = 27.91$; MAE = 0.08; Fig. 7d). According to the relationship between nymphs per stem and TCH, an increase of one unit in the infestation would be associated with a decrease in TCH of 3.4%.
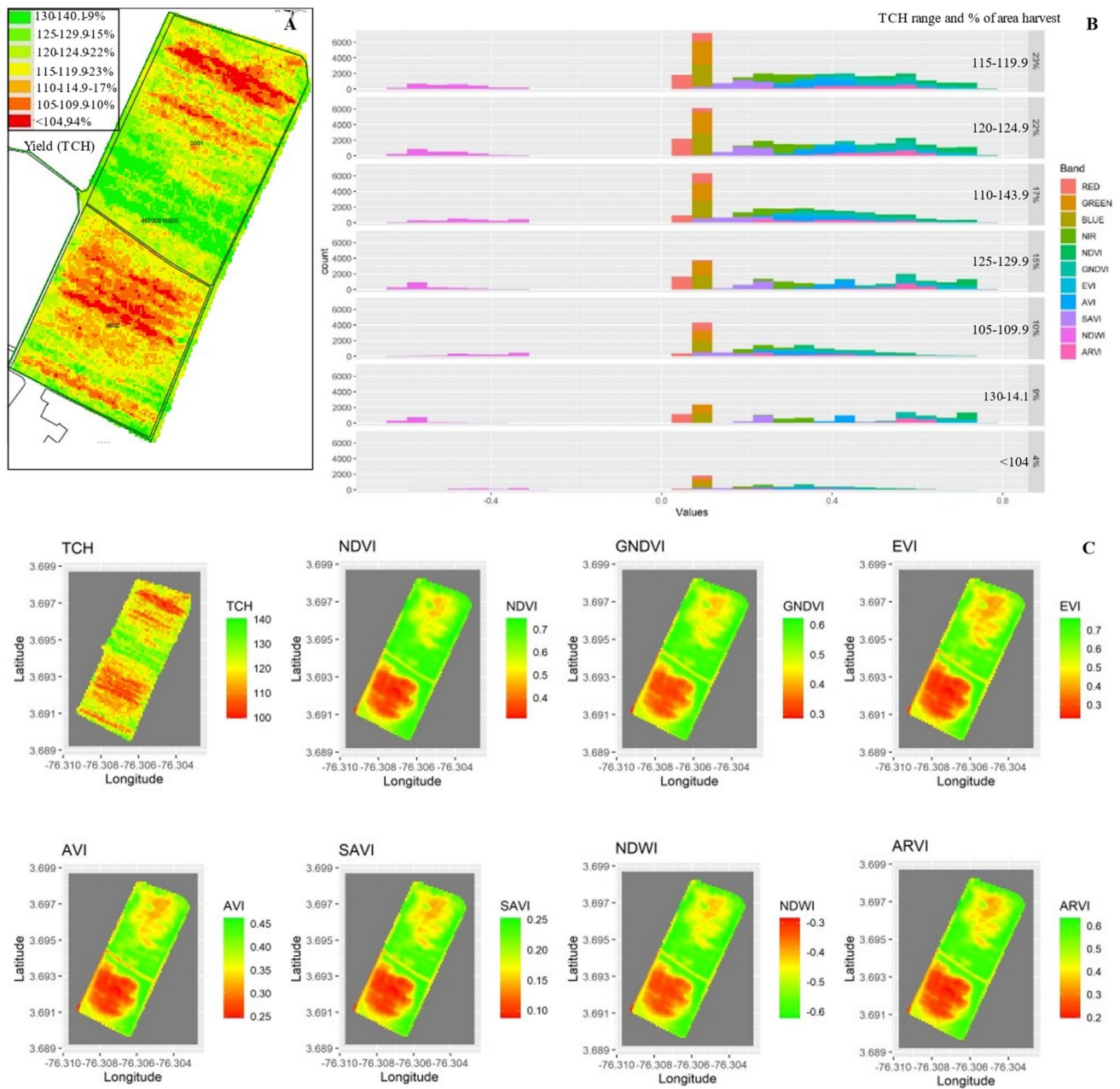
Fig. 5 Visualization of average vegetation indexes and yield in affected plots by infestation by nymphs and adults of *Aeneolamia varia* during a productive cycle. **a** Maps of productivity in terms of Tons of Cane per Hectare (TCH). **b** Percent area with specific range of yield and number of pixels of each band and vegetation index**.** In A and B the productivity map and the value and number of pixels for each index and band evaluated according to the grouping of TCH classes and the% of area, **c** The indexes more effective to inform

about spatial variation on yield (TCH) were NDVI, GNDVI, EVI, AVI, and ARVI based on Lasso regression methods. Normalized difference vegetation index (NDVI); Green normalized difference vegetation index (GNDVI); two-band Enhanced Vegetation Index (EVI2); Ashburn vegetation index (AVI); Soil-adjusted vegetation index (SAVI); Normalized Difference Water Index (NDWI); Atmospheric resistant vegetation (ARVI). TCH: tons of sugarcane per hectare. Lasso: Least Absolute Shrinkage and Selection Operator

Discussion

The results confirmed that the spittlebug *A. varia* is a pest of high economic importance in sugarcane production systems. The infestation levels in this study (means ranging

from 5.1 to 7.4 nymphs per stem and 0.9 to 9.5 adults per stem) were associated with injury that caused losses up to 24.7 tons of cane per hectare, representing an approximately 17% reduction in the expected yield. The foregoing can be aggravated if there is no rapid response in control actions
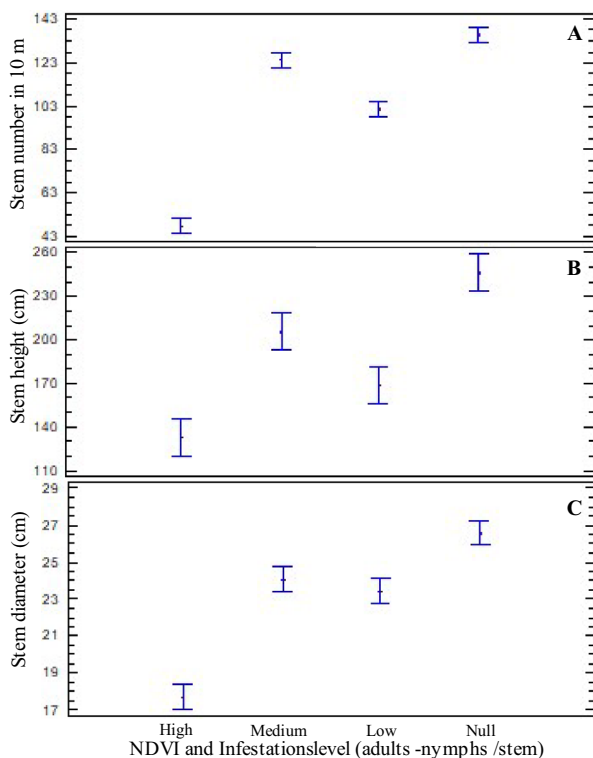
Fig. 6 Monitoring of yield components in affected plots by nymphs and adults of *Aeneolamia varia* in sugarcane at different infestation levels. **a** Mean number of stems in 10 m. **b** Mean stem height (cm). **c** Mean stem diameter (cm). Error bars represent the 95% confidence intervals of the mean, validated by Tukey test. No overlapping of error bars indicates significant differences ($P < 0.05$). IL: Infestation level, where null (0 IL and > 0.58 NDVI), low (0.01–0.14 IL and 0.44–0.57 NDVI), medium (0.141–0.45 IL and 0.34–0.43 NDVI) and high (> 0.451 IL and < 0.33 NDVI). IL was determined as the mean of adults and nymphs per stem. Normalized difference vegetation index (NDVI)

based on sampling of population levels (Table 1). In this case, given the state of the infestation and a delay in detecting the problem, the implementation of management measures was relying mainly on spraying chemical insecticides (Fig. 1a, b).

Spatial analyses using ML tools and satellite images obtained from the GEE platform were effective to assess spittlebug population patterns associated with injury and were also instrumental to estimate final damage (i.e., effects on yield). *A. varia* having an aggregated spatial structure and high space dependence moves relatively quickly from the infested areas, with high impact on neighboring plants evidenced on advanced injury in the surrounding areas (Fig. 1). The latter raises the need for timely detection and early warnings to avoid severe attacks from rapid establishment and population increases (Fig. 1a, b). The space–time dynamics of spittlebugs are highly sensitive to environmental conditions (Figueredo

et al. 2021), which, in the case of the study area, have been found to be highly associated with transitions from dry to rainy periods (Ramirez-Gil et al. 2021), which sets the stage for coupling population dynamics and associated environmental factors to develop early warning systems (Méndez-Vázquez et al. 2019; L. Méndez-Vázquez et al. 2022).

In relation to the above-mentioned, populations of *A. varia* have been found associated with early development of sugarcane, either with those stands with few months after harvest (ratoon) or those during establishment (Vargas et al. 2013), with a greater incidence and relationship with low values in the vegetation indices found in this work (Supplementary Information 1), indicating that the first 6 months of crop development are more sensitive to pest attack, aggravated by the fact that some sugarcane varieties may be more susceptible than others (Cuaran et al. 2012).

As mentioned previously, the injury of spittlebugs in sugarcane is associated with chlorosis, necrosis, portions of reddish brown spots, and in advanced stages death of the stools (Vargas and Gutiérrez 2017). This kind of injury is related to the feeding behavior of both nymphs and adults that after injecting digestive enzymes trigger adverse effects primarily on leaves and then in other tissues (Taliaferro et al. 1967). These leaf lesions cause changes in the coloration of tissues, which can be identified by means of spectral techniques, where the amount of light absorbed by the object and its respective reflectance vary depending on the degree of injury (West et al. 2003; Chaerle et al. 2009; Mahlein et al. 2012). In our results, we identify that bands blue, green, red, and near-infrared, and vegetation index have the capacity to discriminate sugarcane plants with injury by *A. varia*. These bands have been associated with the ability to detect changes in biochemical, color, structural, and leaf density (Vogelmann et al. 1993; Huo et al. 2021).

Negative relationships between variables of infestation and those of crop yield (i.e., TCH and NDVI) corroborated the impact of the pest on production and showed that the negative effect of the infestation is more severe when the infestation is associated with adults than with nymphs. A one-unit increase in adults per stem impacted TCH about five times more than a one-unit increase in nymphs per stem. Notably, this affectation occurred at maximum levels of 1.6 adults per stem, while the nymphs reached maximum levels of 10 nymphs per stem. In studies on Bermuda grass attacked by *Prosapia bicinta* (Say) (Hemiptera: Cercopidae), symptoms of pest attack were only evident when the plants were exposed to adults (Byers and Wells 1966). In sugarcane, attack symptoms are largely caused by adult attacks, 1 to 3 weeks after detection (Fewkes 1969). As greater level of injury was more associated with the presence of adults, it is important to identify how sampling of the nymphal stage can help in preventing the ulterior injury of an increased
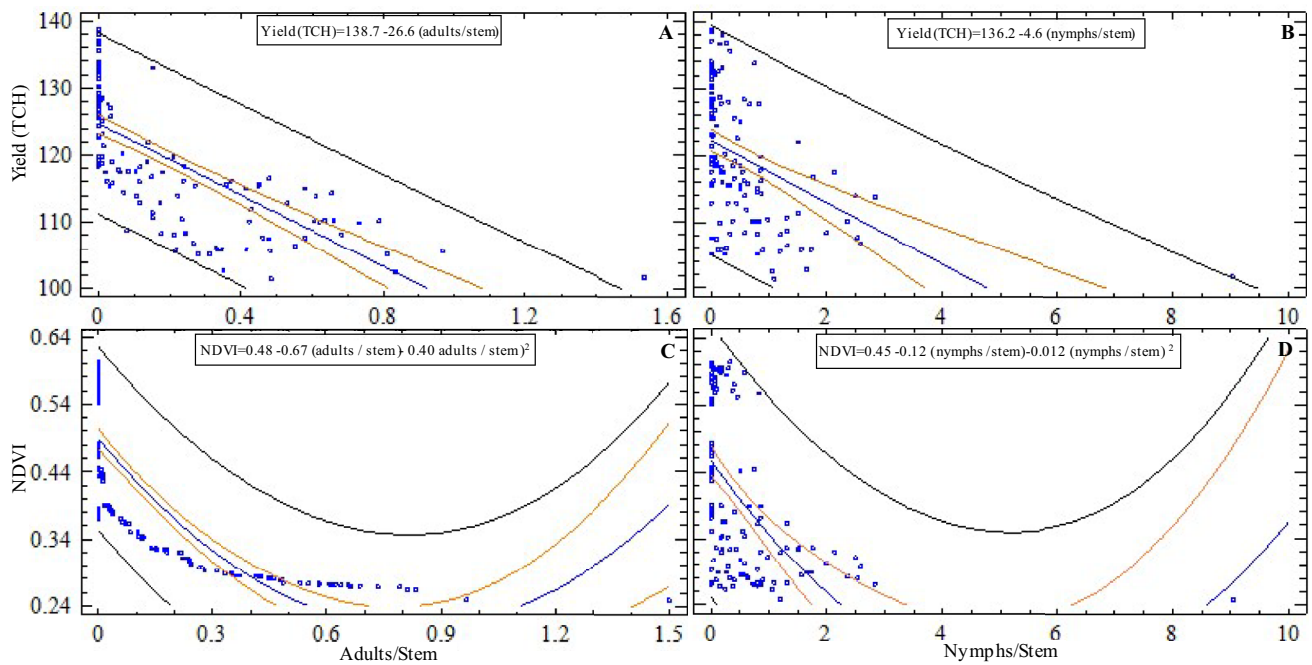
Fig. 7 Fit models for the prediction of biomass (TCH) and NDVI vegetation index in sugar cane associated with infestation by nymphs and adults of *Aeneolamia varia*. A, B, C y D: fit to a linear and second-order quadratic models of biomass production (TCH) and infestation as a function of variables with greater weight determined by the Lasso regression model. Black line: prediction limits.

Orange line: confidence limits. Model statistics metrics: **a** r: −0.61; R^2:51.5%; MAE: 5.6. **b** r: −0.52; R^2:41.4%; MAE: 7.3. **c** r: 0.77; R^2:60.6%; MAE: 4.93. **d** r: 0.75; R^2:60.2%; MAE: 4.94. r coefficient of correlation. R^2: coefficient of determination. *MAE* mean absolute error

population of adults. Under a scenario where the nymphs attack is no noticeable by either plain sight or remote sensing, there would be still need of executing field samplings of nymph populations, especially on those sugarcane fields with history of previous spittlebug attacks or in those plots planted with susceptible varieties, mostly on susceptible crop stages (< 6 months).

According to the above-mentioned, a general approach for IPM (Integrated pest management) of this pest could have remote sensing as complementary tool for injury assessment while filed sampling of both adults and nymphs is sustained in a continuous monitoring program over time (Hutchins 1994). In this ecological aspects need to be considered (Kuno 1991), such as pest's aggregated distribution and its rapid movement in the field, so the detection via remote sensing of areas expressing injury in the field could be an early warning to prevent further increase in the area under attack. The latter while keeping in mind the potential high impact on final biomass production, mainly driven by attacks on early stages of crop development that in the case of the fields surveyed was around 5 months of age.

The approach used on this work represents a combination of promising emerging technologies for plant health monitoring. The GEE platform proved to be suitable for quickly, accurately, and cheaply obtaining a historical series of images for the evaluated batches that, after processing and extracting products such as bands and spectral indices, were incorporated as predictive variables into different ML models and algorithms. The versatility of our approach was due to the fact that, on the one hand, the Sentinel-2 sensor provides constant and uninterrupted multispectral satellite images at high temporal frequency, high spatial resolution and global coverage (Segarra et al. 2020). For its part, GEE efficiently combines numerous remote sensing data sets and processing algorithms, which are easily accessible at multiple spatial scales (Amani et al. 2020). The combination of these emerging technologies simultaneously provides processing infrastructure, storage, algorithms, software, and generation of Internet-based application programming interface (API) and a Web-based interactive development environment, providing unprecedented opportunities to generate effective computational tools for crop monitoring.

This study is illustrating how sampling tools do not only generate information for longitudinal analysis of populations, using elements of descriptive and inferential statistics (Binns et al. 2000), but also can incorporate the space dimension using location and elements of spatial statistics. Additionally, the remote sensing and GEE platform tools used to indirectly detect injury and damage associated with pest infestation is an example of a useful application

associated with the agriculture 4.0 approach (Tyagi 2016; Erazo-Mesa et al. 2022), where advanced technologies, whose objective is the management of information, will enhance the decision-making process (Kamilaris et al. 2017; Erazo-Mesa et al. 2022) with multiple economic benefits and environmental indicators in the production systems where they are implemented (Karmas et al. 2016; Kamilaris et al. 2017; Méndez-Vázquez et al. 2019).

Conclusion

This study aimed to characterize spatio-temporal population dynamics of the spittlebug *A. varia* in commercial sugarcane plots based on the proposal and validation of a digital agriculture approach in which sampling tools, spatiotemporal analyses, satellite sensing big data, and ML analytical techniques are complementary. This approach provided a good approximation for the detection of the infestation level of nymphs and adults, as well as predicting the reduction on yield caused by this pest. The results suggested that the potential impact of *A. varia* infestation varies in sugarcane, generating losses of ~ 17% in biomass, having that further effects on sucrose content warrant additional studies. The GEE platform proved to be an ideal emerging technology to obtain massive satellite data that facilitates characterizations of the spatiotemporal dynamics of *A. varia* attacks on sugarcane. The precise geolocation of the impact of attacks and the magnitude on productivity using the proposed approach generates opportunities to implement emerging management practices that optimize the use of inputs and validate management practices and time control interventions. Future studies should validate the proposed technological approach in similar and different agroecological conditions. Likewise, the agronomic analysis of the results must be complemented with a focus on optimizing crop management practices to prevent or mitigate spittlebug infestations.

Author Contributions JGRG and GV were responsible for conceptualization, investigation, writing—review and editing, and supervision; JGRG and WALR carried out formal analysis; JGRG, WALR, MCF, and GV were involved in writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest. This submission is original to myself and my coauthor, and has not been published previously, nor is it currently under consideration for publication elsewhere. It also complies with all the ethical requirements of the journal. Please do not hesitate to contact us for any questions.

References

Abdel-Rahman, E.M., and F.B. Ahmed. 2008. The application of remote sensing techniques to sugarcane (*Saccharum* spp. hybrid) production: A review of the literature. *International Journal of Remote Sensing* 29: 3753–3767. https://doi.org/10.1080/01431160701874603.

Aguilar-Rivera, Noé. 2019. A framework for the analysis of socio-economic and geographic sugarcane agro industry sustainability. *Socio-Economic Planning Sciences* 66: 149–160. https://doi.org/10.1016/j.seps.2018.07.006.

Allsopp, P.G., and R.M. Bull. 1989. Spatial patterns and sequential sampling plans for melolonthine larvae (Coleoptera: Scarabaeidae) in southern Queensland sugarcane. *Bulletin of Entomological Research* 79: 251–258.

Amani, M., M. Kakooei, A. Moghimi, A. Ghorbanian, B. Ranjgar, S. Mahdavi, A. Davidson, T. Fisette, P. Rollin, and B. Brisco. 2020. Application of google earth engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in Canada. *Remote Sensing* 12: 3561.

Arbogast, Richard T., Paul E. Kendra, Richard W. Mankin, and Jeffrey E. McGovern. 2000. Monitoring insect pests in retail stores by trapping and spatial analysis. *Journal of Economic Entomology* 93: 1531–1542.

Arias, C., A. Martínez. Valencia, JG Morales. Osorio, and Osorio, and J. G. Ramírez-Gil. 2019. Spatial analysis of presence, injury, and economic impact of the Melolonthidae (Coleoptera: Scarabaeoidea) complex in avocado crops. *Neotropical Entomology* 48: 583–593.

Asocaña. 2020. Informe anual de Asocaña con aspectos generales del Sector Agroindustrial de la Caña de Colombia 2019–2020 y Anexos estadístico. Asocaña.

Baddeley, Adrian, Ege Rubak, and Rolf Turner. 2015. *Spatial point patterns: Methodology and applications with R*. London: CRC Press.

Bekkar, Mohamed, Hassiba Kheliouane Djemaa, and Taklit Akrouf Alitouche. 2013. Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications* 3: 27–38.

Binns, Michael R., Jan P. Nyrop, Wopke van der Werf, and Wopke Werf. 2000. *Sampling and monitoring in crop protection: the theoretical basis for developing practical decision guides*. Cabi.

Bivand, Roger, C Rundel, E Pebesma, R Stuetz, K Hufthammer, P Giraudoux, M Davis, and S Santilli. 2020. *rgeos: Interface to Geometry Engine-Open Source ('GEOS')*. Comprehensive R Archive Network (CRAN).

Breiman, Leo. 1996. Bagging predictors. *Machine Learning* 24: 123–140. https://doi.org/10.1007/BF00058655.

Brenner, Richard J., Dana A. Focks, Richard T. Arbogast, David K. Weaver, and Dennis Shuman. 1998. Practical use of spatial analysis in precision targeting for integrated pest management. *American Entomologist* 44: 79–102.

Brownlee, Jason. 2016. *Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras*. Machine Learning Mastery.

Byers, R.A., and Homer D. Wells. 1966. Phytotoxemia of coastal Bermudagrass caused by the two-lined spittlebug, *Prosapia bicincta* (Homoptera: Cercopidae)1. *Annals of the Entomological Society of America* 59: 1067–1071. https://doi.org/10.1093/aesa/59.6.1067.

Carbonell González, J. 2011. *Zonificación agroecológica para el cultivo de la caña de azúcar en el valle del río Cauca (cuarta aproximación): principios metodológicos y aplicaciones*. cenicaña (Colombia).

Chaerle, Laury, Sándor. Lenk, Ilkka Leinonen, Hamlyn G. Jones, Dominique Van Der Straeten, and Claus Buschmann. 2009. Multisensor plant imaging: Towards the development of a stress-catalogue. *Biotechnology Journal* 4: 1152–1167. https://doi.org/10.1002/biot.200800242.

Cherry, R.H. 1984. Spatial distribution of white grubs (Coleoptera: Scarabaeidae) in Florida sugarcane. *Journal of Economic Entomology* 77: 1341–1343.

Cortes, Corinna, and Mehryar Mohri. 2003. AUC optimization vs. error rate minimization. In *Proceedings of the 16th international conference on neural information processing systems*, 313–320. NIPS'03. Whistler: MIT Press.

Cressie, Noel. 1991. *Statistics for spatial data*. New York: Wiley.

Cuaran, V.L., Ulises Castro Valderrama, Alex Enrique Bustillo. Pardey, Nora Cristina Mesa. Cobo, Gerson Darío Ramírez. Sánchez, Carlos Arturo Moreno. Gil, and Luis Antonio Gómez. Laverde. 2012. Método para evaluar el daño de los salivazos (Hemiptera: Cercopidae) sobre caña de azúcar, Saccharum spp. *Revista Colombiana De Entomología* 38: 171–176.

Deleon, Leonel, Michael J. Brewer, Isaac L. Esquivel, and Jonda Halcomb. 2017. Use of a geographic information system to produce pest monitoring maps for south Texas cotton and sorghum land managers. *Crop Protection* 101: 50–57.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. Least angle regression. *Annals of Statistics* 32: 407–451.

Erazo-Mesa, Edwin, Andrés Echeverri-Sánchez, and Joaquin Guillermo Ramírez-Gil. 2022. Advances in Hass avocado irrigation scheduling under digital agriculture approach. *Revista Colombiana De Ciencias Hortícolas* 16: e13456–e13456. https://doi.org/10.17584/rcch.2022v16i1.13456.

Ferguson, Andrew W., Zdisław Klukowski, Barbara Walczak, Suzanne J. Clark, Moira A. Mugglestone, Joe N. Perry, and Ingrid H. Williams. 2003. Spatial distribution of pest insects in oilseed rape: implications for integrated pest management. *Agriculture, Ecosystems and Environment* 95: 509–521.

Fewkes, D.W. 1969. The biology of sugar cane froghoppers. In *Pests of sugar cane*, ed. J.R. Williams, J.R. Metcalfe, R.W. Mungomery, and R. Mathes, 283–307. Amsterdam: Elsevier Publishing Company.

Figueredo, Luis, Adriana Villa-Murillo, Yelitza Colmenarez, and Carlos Vásquez. 2021. A hybrid artificial intelligence model for *Aeneolamia varia* (Hemiptera: Cercopidae) populations in sugarcane crops. *Journal of Insect Science* 21: 11. https://doi.org/10.1093/jisesa/ieab017.

Fox, John, Sanford Weisberg, Brad Price, Daniel Adler, Douglas Bates, Gabriel Baud-Bovy, Ben Bolker, et al. 2020. *Car: Companion to Applied Regression* (version 3.0–10).

Friedman, Jerome, Trevor Hastie, Rob Tibshirani, Balasubramanian Narasimhan, Kenneth Tay, Noah Simon, and Junyang Qian. 2021. Package 'glmnet.' *CRAN R Repository*.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1.

Genc, H., L. Genc, H. Turhan, S.E. Smith, and J.L. Nation. 2008. Vegetation indices as indicators of damage by the sunn pest (Hemiptera: Scutelleridae) to field grown wheat. *African Journal of Biotechnology*. https://doi.org/10.4314/ajb.v7i2.58347.

Ghazvinei, Pezhman Taherei, Hossein Hassanpour Darvishi, Amir Mosavi, Khamaruzaman Bin Wan. Yusof, Meysam Alizamir, Shahaboddin Shamshirband, and Kwok-wing Chau. 2018. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Engineering Applications of Computational Fluid Mechanics* 12: 738–749. https://doi.org/10.1080/19942060.2018.1526119.

Gigot, Christophe. 2018. *Analyzing plant disease epidemics with R package epiphy*.

Hasmadi, Mohd, H.Z. Pakhriazad, and M.F. Shahrin. 2009. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia-Malaysian Journal of Society and Space* 5: 1–10.

Hastie, T. 2020. *gam: Generalized Additive Models* (version 1.20). Comprehensive R Archive Network (CRAN).

Hatfield, P.L., and P.J. Pinter Jr. 1993. Remote sensing for crop protection. *Crop Protection* 12: 403–413.

He, Yong, Hong Zeng, Yangyang Fan, Shuaisheng Ji, and Wu. Jianjian. 2019. Application of deep learning in integrated pest management: A real-time system for detection and diagnosis of oilseed rape pests. *Mobile Information Systems* 2019: 1–14.

Henao-Rojas, Juan Camilo, María Gladis. Rosero-Alpala, Carolina Ortiz-Muñoz, Carlos Enrique Velásquez-Arroyo, William Alfonso Leon-Rueda, and Joaquín Guillermo. Ramírez-Gil. 2021. Machine learning applications and optimization of clustering methods improve the selection of descriptors in blackberry Germplasm Banks. *Plants*. https://doi.org/10.3390/plants10020247.

Hijman, R. J. 2016. Package 'raster.'

Huo, Langning, Henrik Jan Persson, and Eva Lindberg. 2021. Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red and SWIR (NDRS). *Remote Sensing of Environment* 255: 112240. https://doi.org/10.1016/j.rse.2020.112240.

Hutchins, Scott H. 1994. Techniques for sampling arthropods in integrated pest management. In *Handbook of sampling methods for arthropods in agriculture*. CRC Press.

Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. 2017. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture* 143: 23–37. https://doi.org/10.1016/j.compag.2017.09.037.

Karmas, A., A. Tzotsos, and K. Karantzalos. 2016. *Geospatial big data for environmental and agricultural applications*. Berlin: Springer.

Khdery, Ghada A. 2021. Remote sensing technology and its applications in plant pathology. In *Emerging trends in plant pathology*, 683–701. Springer.

Kuhn, Max, Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, et al. 2020. *caret: Classification and regression training* (version 6.0–85).

Kuno, E. 1991. Sampling and analysis of insect populations. *Annual Review of Entomology* 36: 285–304. https://doi.org/10.1146/annurev.en.36.010191.001441.

Li, Baohua, Laurence V. Madden, and Xu. Xiangming. 2012. Spatial analysis by distance indices: An alternative local clustering index for studying spatial patterns. *Methods in Ecology and Evolution* 3: 368–377.

Liaw, A, and M Wiener. 2018. *Randomforest: Breiman and Cutler's random forests for classification and regression* (version 4.6–14).

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2: 18–22.

Mahlein, Anne-Katrin., Erich-Christian. Oerke, Ulrike Steiner, and Heinz-Wilhelm. Dehne. 2012. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology* 133: 197–209. https://doi.org/10.1007/s10658-011-9878-z.

Méndez-Vázquez, L., R. Lasa-Covarrubias, S. Cerdeira-Estrada, and A. Lira-Noriega. 2022. Using simulated pest models and biological clustering validation to improve zoning methods in site-specific pest management. *Applied Sciences* 12: 1900. https://doi.org/10.3390/app12041900.

Méndez-Vázquez, L. Josué., Andrés Lira-Noriega, Rodrigo Lasa-Covarrubias, and Sergio Cerdeira-Estrada. 2019. Delineation of site-specific management zones for pest control purposes: Exploring precision agriculture and species distribution modeling approaches. *Computers and Electronics in Agriculture* 167: 105101. https://doi.org/10.1016/j.compag.2019.105101.

Militante, S. V., B. D. Gerardo, and R. P. Medina. 2019. Sugarcane Disease Recognition using Deep Learning. In *2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE)*, 575–578. https://doi.org/10.1109/ECICE47484.2019.8942690.

Mondal, Pinki. 2011. Quantifying surface gradients with a 2-band Enhanced Vegetation Index (EVI2). *Ecological Indicators* 11: 918–924. https://doi.org/10.1016/j.ecolind.2010.10.006.

Moran, P.A.P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B (methodological)* 10: 243–251.

Mueller, A., and S. Guido. 2016. *Introduction to machine learning with python: A guide for data scientists*. Dawn Sachanafelt: O'Reilly Media.

Natarajan, Rajathi, Jayashree Subramanian, and Elpiniki I. Papageorgiou. 2016. Hybrid learning of fuzzy cognitive maps for sugarcane yield classification. *Computers and Electronics in Agriculture* 127: 147–157. https://doi.org/10.1016/j.compag.2016.05.016.

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26: 217–222. https://doi.org/10.1080/01431160412331269698.

Paluszynska, Aleksandra, Przemyslaw Biecek, Yue Jiang [aut, and cre. 2019. *randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance* (version 0.10.0).

Pardede, Hilman F., Endang Suryawati, Dikdik Krisnandi, R. Sandra Yuwana, and Vicky Zilvan. 2020. Machine learning based plant diseases detection: A review. In *2020 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET)*, 212–217. IEEE.

Partel, Victor, Leon Nunes, Phil Stansly, and Yiannis Ampatzidis. 2019. Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence. *Computers and Electronics in Agriculture* 162: 328–336.

Pavlu, Franz Arthur, and José Paulo. Molin. 2016. A sampling plan and spatial distribution for site-specific control of Sphenophorus levis in sugarcane. *Acta Scientiarum. Agronomy* 38: 279–287.

Pebesma, E. 2018. Simple features for R: Standardized support for spatial vector data. *The R Journal*. https://doi.org/10.32614/RJ-2018-009.

Perry, Joe N., L. Winder, J. M. Holland, and R. D. Alston. 1999b. Red–blue plots for detecting clusters in count data. *Ecology Letters*. Wiley Online Library.

Perry, Joe N. 1995. Spatial analysis by distance indices. *Journal of Animal Ecology* 64: 303–314.

Perry, J.N., L. Winder, J.M. Holland, and R.D. Alston. 1999a. Red–blue plots for detecting clusters in count data. *Ecology Letters* 2: 106–113. https://doi.org/10.1046/j.1461-0248.1999.22057.x.

QGIS Development. 2016. QGIS system. *Open source geospatial foundation project*.

R Development Core Team. 2022. R: The R Project for Statistical Computing. 2022. *R foundation for statistical computing*. Vienna: Austria.

Ramirez-Gil, JG, German Vargas, Hector Chica-Ramirez, Alberto Roldan-Gonzalez, and A J Peña. 2021. Variación temporal de la lluvia como modulador de la dinámica poblacional del salivazo Aeneolamia varia (Hemiptera: Cercopidae) en sistemas de producción de caña de azúcar. In *Congreso Internacional de Variabilidad y Cambio Climático*, 45–46. Bogotá: Universidad Nacional de Colombia.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, Markus Müller, Stefan Siegert (Fast DeLong code), and Matthias Doering (Hand & Till Multiclass). 2021. *pROC: Display and analyze ROC curves* (version 1.17.0.1).

Rustia, Dan Jeric, Chien Erh Arcega, Jui-Yung Chung. Lin, Yi.-Ji. Zhuang, Ju-Chun. Hsu, and Ta.-Te. Lin. 2020. Application of an image and environmental sensor network for automated greenhouse insect pest monitoring. *Journal of Asia-Pacific Entomology* 23: 17–28.

Salazar, J., and L. Proaño. 1989. Pérdidas ocasionadas por la candelilla de la caña de azúcar (Aeneolamia varia) en el área de influencia del central río Turbio: Estudio comparativo de las zafras 84/85 y 85/86. *Caña De Azúcar* 7: 49–54.

Sarparast, Mehdi. 2018. *LSRS: Land Surface Remote Sensing* (version 0.2.0).

Schexnayder, H.P., Jr., T.E. Reagan, and D.R. Ring. 2001. Sampling for the sugarcane borer (Lepidoptera: Crambidae) on sugarcane in Louisiana. *Journal of Economic Entomology* 94: 766–771.

Segarra, J., M.L. Buchaillot, J.L. Araus, and S.C. Kefauver. 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy* 10: 641.

Souza, M.F.D., L.R.D. Amaral, S.R.D.M. Oliveira, M.A.N. Coutinho, and C.F. Netto. 2020. Spectral differentiation of sugarcane from weeds. *Biosystems Engineering* 190: 41–46. https://doi.org/10.1016/j.biosystemseng.2019.11.023.

Sudbrink, Donald L., Steven J. Thomson, Reginald S. Fletcher, F. Aubrey Harris, Patrick J. English, and James T. Robbins. 2015. Remote sensing of selected winter and spring host plants of tarnished plant bug (Heteroptera: Miridae) and herbicide use strategies as a management tactic. *American Journal of Plant Sciences* 06: 1313. https://doi.org/10.4236/ajps.2015.68131.

Taliaferro, C.M., R.A. Byers, and G.W. Burton. 1967. Effects of spittlebug injury on root production and sod reserves of coastal Bermudagrass1. *Agronomy Journal* 59: 530–532. https://doi.org/10.2134/agronj1967.00021962005900060013x.

Tyagi, Avinash C. 2016. Towards a second green revolution. *Irrigation and Drainage* 65: 388–389. https://doi.org/10.1002/ird.2076.

Vargas, G, U Castro, Y Granobles, and G Ramirez. 2013. Biología y perspectivas de manejo de los salivazos Mahanarva bipars y Aeneolamia varia (Hemiptera: Cercopidae) en caña de azúcar. In *XXXX congreso de la Sociedad Colombiana de Entomología,*

Socolen, 422–435. Bogotá: Sociedad Colombiana de Entomología, Socolen.

Vargas, G., and Y. Gutiérrez. 2017. Manejo de las poblaciones de Aeneolamia varia. *Carta Informativa Cenicaña. No.* 1: 18–19.

Vogelmann, J.E., B.N. Rock, and D.M. Moss. 1993. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing* 14: 1563–1575. https://doi.org/10.1080/01431169308953986.

Wang, Ming, Zhengjia Liu, Muhammad Hasan Ali. Baig, Yongsheng Wang, Yurui Li, and Yuanyan Chen. 2019. Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. *Land Use Policy* 88: 104190. https://doi.org/10.1016/j.landusepol.2019.104190.

West, Jonathan S., Cedric Bravo, Roberto Oberti, Dimitri Lemaire, Dimitrios Moshou, and H. Alastair McCartney. 2003. The potential of optical canopy measurement for targeted control of field crop diseases. *Annual Review of Phytopathology* 41: 593–614. https://doi.org/10.1146/annurev.phyto.41.121702.103726.

Wu, Jingan, Liupeng Lin, Tongwen Li, Qing Cheng, Chi Zhang, and Huanfeng Shen. 2022. Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network. *International Journal of Applied Earth Observation and Geoinformation* 108: 102738. https://doi.org/10.1016/j.jag.2022.102738.

Yang, Chwen-Ming., Ching-Huan. Cheng, and Rong-Kuen. Chen. 2007. Changes in spectral characteristics of rice canopy infested with brown Planthopper and Leaffolder. *Crop Science* 47: 329–335. https://doi.org/10.2135/cropsci2006.05.0335.

Yang, Z., M.N. Rao, N.C. Elliott, S.D. Kindler, and T.W. Popham. 2005. Using ground-based multispectral radiometry to detect stress in wheat caused by greenbug (Homoptera: Aphididae) infestation. *Computers and Electronics in Agriculture* 47: 121–135. https://doi.org/10.1016/j.compag.2004.11.018.