Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v44e20230148/2024

CLASSIFICATION OF THE OCCURRENCE OF BROADLEAF WEEDS IN NARROW-LEAF CROPS

Cenneya L. Martins[1*], Agda L. G. Oliveira[1], Isabella A. da Cunha[1], Henrique Oldoni[2], Juliana C. Pereira[1], Lucas R. do Amaral[1]

[1*]Corresponding author. Universidade Estadual de Campinas - UNICAMP, School of Agricultural Engineering - FEAGRI/Campinas - SP, Brazil.
E-mail: cenneya.martins@feagri.unicamp.br | ORCID ID: https://orcid.org/0000-0002-4585-2739

KEYWORDS

precision agriculture, machine learning, remote sensing, weed management.

ABSTRACT

Considering the spectral differences between broadleaf weeds and narrow-leaf crops and the influence of terrain and soil variables on weed infestations, integrating such information into a machine-learning algorithm can lead to accurate weed maps. Therefore, we aim to evaluate the effectiveness of these variables in classifying the occurrence of broadleaf weeds in narrow-leaf crops. Weed data was collected at georeferenced points across two areas covering 200 ha (pasture) and 106 ha (sorghum), creating classes 0 (absence) and 1 (presence). For each sample point, we obtained 11 variables: soil clay content, cation exchange capacity, soil organic matter, terrain elevation, slope, NDVI, EVI, CIgreen, BGND (derived from PlanetScope images), and spatial information (X and Y coordinates). These variables were used as predictors of broadleaf weed presence and absence in the Random Forest classification algorithm. The presence and absence of broadleaf weeds were correctly classified in 84% and 74% of all predictions in the test sample sets for pasture and sorghum areas, respectively. This strategy represents an efficient way to map and manage the occurrence of broadleaf weeds in narrow-leaf crops.

INTRODUCTION

Manual weed sampling for mapping and site-specific management is a costly and inefficient process, especially to identify infestations in extensive areas. As a result, traditional weed control is carried out over the entire area based on the average infestation rate. However, weeds exhibit spatial variability and typically occur in aggregated patterns (Martín et al., 2015), which favors site-specific management. Therefore, demand for tools that enable the mapping of weed occurrence and allow for localized control, leading to agronomic, economic, and environmental benefits, is growing.

Using imagery collected by sensors onboard remotely piloted aircraft (RPA), popularly known as drones, has become a strategy for digital mapping and weed management (Hunter et al., 2020). An example is the utilization of high spatial and spectral resolution sensors, combined with machine learning, to generate vegetation maps, distinguishing between crops and weeds. However, this mapping type has specific characteristics that make obtaining the desired products challenging. These challenges include additional image processing of segmented images, complex information fusion techniques, high computational costs, ground sample distance (GSD) restrictions, and resolution loss due to image quality reduction (Sa et al., 2018). Additionally, to achieve quality imaging, RPA flights must be conducted when there is no shading in the inter-row spaces, usually with wind speeds not exceeding 36 km/h and under clear sky conditions.

Satellite images can be more easily obtained compared with RPA images. However, the ability to spectrally differentiate between vegetation and weeds can

[1] Universidade Estadual de Campinas - UNICAMP, School of Agricultural Engineering - FEAGRI/Campinas - SP, Brazil.
[2] Universidade Estadual de Campinas - UNICAMP, Interdisciplinary Center of Energy Planning - NIPE/Campinas - SP, Brazil.

Area Editor: Teresa Cristina Tarlé Pissarra
Received in: 10-31-2023
Accepted in: 2-7-2024

be affected by the spectral and spatial resolution of these sensors due to the spectral similarities between weeds and crops (Lamb & Brown, 2001). Additionally, soil background effects impact weed differentiation after crop emergence (Thorp & Tian, 2004).

As a weed management strategy, monocotyledonous and dicotyledonous weeds, due to their morphological and physiological variations, are commonly categorized as "narrow-leaf" and "broadleaf," respectively. In this regard, broadleaf plants generally exhibit spectral and architectural characteristics that clearly distinguish them from narrow-leaf plants, favoring their differentiation by remote sensing techniques (Gitelson et al., 2005). In this context, Souza et al. (2020) observed significant spectral differences between sugarcane (monocotyledonous) and weeds, especially those in the dicotyledonous group. This can also be considered for other narrow-leaf crops, such as pasture and sorghum. Note that vegetation indices (VIs), which enhance spectral features of interest, provide a valuable alternative in remote sensing for agricultural monitoring.

In digital soil mapping, an alternative aimed at increasing the reliability of spatial predictions and reducing sampling costs is the use of variables associated with specific soil attributes as predictor variables (Pusch et al., 2023). For example, topographic features, climate, vegetation indices, and parent material of the soil (Wadoux et al., 2020) can be employed in creating maps that represent soil attributes like organic matter and pH (Szatmári et al., 2019). Considering that the spatial distribution of weeds can vary due to dispersion factors (wind, water flow, and others), soil-related factors, and the soil seed bank (Metcalfe et al., 2019; Nordmeyer, 2006), environmental variables can similarly be used for weed mapping.

Integrating multiple sources of information associated with weed infestations, such as topographic data, soil data, and spectral vegetation data as predictors of the presence and absence of broadleaf weeds in narrow-leaf crops, can meet the need for sensors with high spatial resolution and allow for the generation of maps with a low error rate for weed management in precision agriculture. Therefore, we developed classification models within a machine learning algorithm using the mentioned variables as indicators of the occurrence of broadleaf weeds in areas where pasture and sorghum are cultivated. Our goal is to assess the effectiveness of these variables in classifying the occurrence of broadleaf weeds in narrow-leaf crops.

MATERIAL AND METHODS

Experimental Areas

The research was conducted in 2019 and 2023 in two agricultural commercial production areas in the state of São Paulo, Brazil (Figure 1). Area A, situated in the municipality of Caiuá (21°38'15.5"S, 51°54'44.9"W), spans 200 ha in an integrated crop-livestock management system, with pasture composed of a mix of *Brachiaria* and millet, preceded by soybean cultivation. The pasture was seeded from March 31, 2019, to April 6, 2019. Area B, located in the municipality of Cosmópolis (22°41'56.7"S, 47°10'32.5"W), covers 106 ha in a crop succession system, where sorghum was sown on March 20, 2023, preceding the soybean crop. Both areas are managed in a no-tillage system and the fertilization and pesticide application schedules followed the grower standards.

The climate in Area A is classified, according to Köppen, as tropical with a dry winter (Aw), with an average annual temperature of 24 °C and precipitation of 1,400 mm (Alvares et al., 2013). The terrain features gently undulating topography, and the soils are predominantly categorized as Eutrophic Clayey Latosol with medium texture (16 to 25% clay). In Area B, the climate is classified as tropical with a hot summer (Cfa), according to Köppen, with an average annual temperature of approximately 20 °C and annual precipitation of 1,600 mm (Alvares et al., 2013). The terrain is gently undulating, and the soil is predominantly classified as typical mesic latosol with clayey texture (36 – 60% clay).

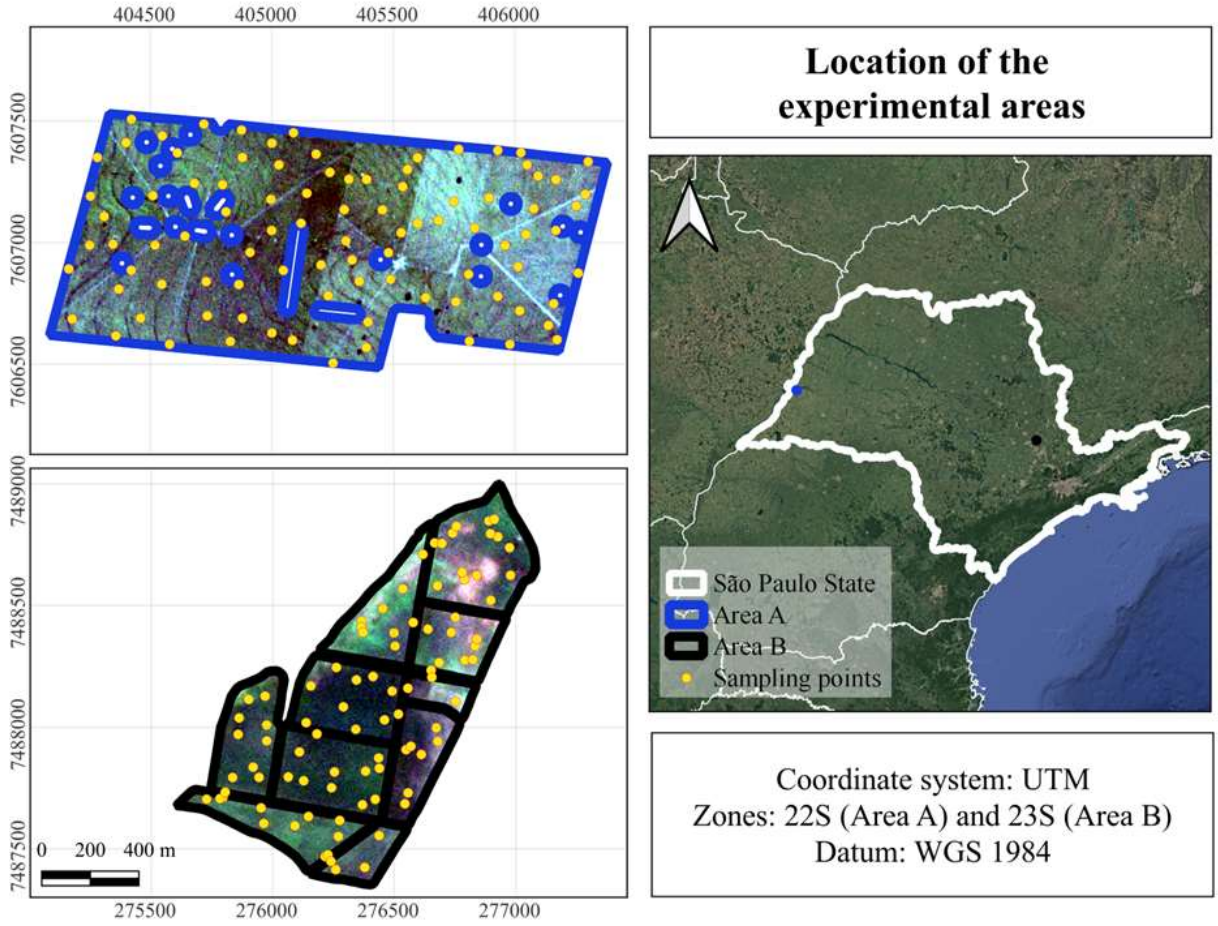

FIGURE 1. Location of the experimental areas and sampling points. Area A – Caiuá, PlanetScope image 07/05/2019, true color composite (red, green, and blue). Area B – Cosmópolis, PlanetScope image 06/05/2023, true color composite (red, green, and blue). São Paulo database: IBGE and Google Earth, 2023.

Predictor Variables

In the classification models, we used 11 predictor variables that, according to the literature, could be related to weed infestations. These include three soil variables, two topographic variables, four vegetation indices, and spatial information (X and Y coordinates).

a) Soil Variables

We conducted soil sampling at 0 to 0.2 m depth to analyze soil fertility and texture in the areas using an automated drill-type soil sampler (Model Trail Tech 500) attached to a quad bike. Each composite sample comprised four individual samples collected within a 3 m radius over up to three subsequent days. Area A had 100 sampling points and Area B had 89 (Figure 1). In Area A, sampling points were randomly stratified throughout the field. In Area B, the points were allocated according to a multi-objective optimized spatial sampling plan (SPSANN – optimization of spatial samples via simulated annealing), as used in a study for predicting soil variables (for more details, refer to Pusch et al., 2023). Among the results of soil analyses, we used clay content (g/kg), cation exchange capacity (CEC – mmol/dm^3), and soil organic matter (SOM – g/kg) as predictor variables in the classification algorithm. Maps of these soil attributes were constructed using geostatistical interpolation with ordinary kriging. The choice of the best semivariogram model (spherical, exponential, or Gaussian) was based on the optimal values of root mean squared error (RMSE) and coefficient of determination (R2), calculated by leave-one-out cross-validation.

b) Topographical characteristics

We collected elevation data densely across the areas using a global navigation satellite system (GNSS) receiver with differential correction, model Starfire 7000-SF1, attached to the harvester. The GNSS receiver allows continuous data collection along the machine's path, resulting in dense terrain coverage, and the differential correction compensates for potential systematic errors in coordinates. We interpolated this data using ordinary kriging, similar to the approach used for soil data. However, due to the high volume of data, we implemented the 10-fold cross-validation method to validate the interpolation model. From the elevation map, we calculated the terrain slope (rad) using the RSAGA package in the R software (Brenning et al., 2022).

c) Vegetation indices

We obtained a PlanetScope system image for both areas with a spatial resolution of 3 m close to the sampling date, without cloud interference. From the individual spectral bands of the images, we calculated the vegetation indices (VIs) NDVI and EVI, widely used in agricultural studies; CIgreen, developed for spectral differentiation of species; and BGND, which uses bands in the visible region (Table 1).

TABLE 1. Vegetation indices.

Vegetation indices	Names	Formulas	References
NDVI	*Normalized difference vegetation index*	$\frac{NIR - R}{NIR + RED}$	(Rouse et al., 1973)
EVI	*Enhanced vegetation index*	$G\frac{NIR -}{NIR + C1\,R - C2\,B + L}$	(Huete et al., 1999)
CI_{green}	*Chlorophyll index – Green*	$\frac{NIR}{G} - 1$	(Gitelson et al., 2005)
BGND	*Blue Green Normalized Difference*	$\frac{G - B}{G + B}$	-

*Spectral bands: NIR = Near-Infrared; R = Red; RE = RedEdge; G = Green; B = Blue.
*EVI: C1 and C2 = coefficients for aerosol resistance (C1 = 6; C2 = 7.5).

Weed Sampling

Weed sampling in the areas was co-located with soil sampling points (Figure 1). In Area A, we conducted weed sampling in July 2019. At the time of sampling, during a period of low water availability, the pasture was underdeveloped. Emerged broadleaf weed species at the sampling point were counted and identified. The presence and absence of broadleaf weeds resulted in a binary data column: 0 (absence) and 1 (presence). We considered weeds presence when at least one plant was present in a 1 m² area.

In Area B, we conducted weed sampling in June 2023 during the flowering stage of sorghum. At each sampling point, the absence (0) or presence (1) of broadleaf weeds was visually classified when they occupied at least 50% of the sampling point area (3 × 3 m area, considering the size of the pixels in the PlanetScope images used in this study). Subsequently, weed data were co-located with the data from all predictor variables.

Analysis

a) Class Balancing

To avoid model bias toward the majority class, we balanced the class with the synthetic minority oversampling technique (SMOTE). This technique generates synthetic examples from the minority class by introducing new instances that are weighted combinations of existing observations (Chawla et al., 2002). Area A had 100 points before balancing, and only 20% of the data belonged to class 0. After balancing, we obtained 126 points, with 50% of the data in each class. Out of 89 sample points in Area B, 48% of the data belonged to class 0, and balancing was unnecessary.

b) Classification of occurrence of broadleaf weeds

Using the random forest (RF) machine learning algorithm, we employed the variables to classify broadleaf weeds in both areas separately. The data was randomly divided into 70% training and 30% testing sets. All variables were used as predictors for classes 0 and 1, assigned to the absence and presence of broadleaf weeds (target variable). To choose the best hyperparameter values, we performed an optimization using the random search method (1000 iterations) with the number of trees tested between 100 and 3000 (ntree), the number of variables selected at each split from 1 to 14 (mtry), and the minimum number of terminations from 2 to 30 (nodesize). The hyperparameters were evaluated with 10-fold cross-validation on the training set, considering the highest accuracy value as the best choice. Subsequently, confusion matrices were generated with the test data (30% of balanced data). This includes 38 samples in Area A and 26 in Area B, with both areas having 50% of these data belonging to class 0 and 50% to class 1. From the data in each confusion matrix, we calculated the metrics: accuracy, precision, specificity, recall, and F1 score. Accuracy is a general metric that measures the proportion of the model's correct predictions from the total predictions. Precision measures the proportion of correct positive predictions from the total positive predictions classified by the model (Equation 2). Specificity measures the proportion of true negatives from the total real negative cases (Equation 3). Recall or sensitivity measures the proportion of correct positive predictions made by the model from the total real positive cases (Equation 4). The F1 Score is a metric that combines precision and recall into a single score and indicates whether these metrics are balanced (Equation 5).

$$Accuracy = \frac{(TN+\quad)}{(TN+FP+FN+T\ )} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (3)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

$$F1\ Score = 2\frac{(Precision \times Recall)}{(Precision+Recal\ )} \quad (5)$$

In which:

TN – True negative;

TP – True positive;

FN – False negative,

FP – False positive.

Based on the classifications, we also ranked the variables and selected those that presented at least 70% importance for the model. After the selection step, we again classified broadleaf weeds using only the selected variables, separately for each study area. The performances of the model constructed with all predictor variables and the model with the selected variables were evaluated by applying each model to the test set. When the variable selection worsened the classifications, we applied the model to the maps of all variables to predict the broadleaf weed map. When the selection allowed for better classifications or the classifications remained the same as those obtained with all variables, we constructed the map of the presence and absence of broadleaf weeds by applying the model to the maps of the selected variables. For these analyses, we used the mlr package in the R software (Bischl et al., 2016).

RESULTS AND DISCUSSION

In Area A, the broadleaf weed species found were: *Macroptilium lathyroides*, *Crotalaria incanum*, *Sida* sp., *Calopogonium mucunoides*, *Macroptilum atropurpureum*, *Senna occidentalis*, *Commelina benghalensis*, *Macroptilium lathyroides*, *Gomphrena globosa*, *Desmodium tortuosum*, *Amaranthus deflexus*, *Physalis angulata*, *Alysicarpus vaginalis*, and *Glycine max* (soybeans from the previous harvest).

In Area B, the broadleaf weed species found were: *Bidens pilosa*, *Commelia benghalensis L., Amaranthus* sp., *Ipomoea* sp., *Alternanthera sessilis* L., and *Solanum americanum*. Note that the species *Commelina benghalensis* is often categorized as a broadleaf weed, due to its morphological characteristics that resemble this group. However, the species is a monocotyledonous plant. To facilitate the distinction in the context of pasture and sorghum, we decided to include it as a broadleaf plant.

The strategy combining various sources of information as predictor variables in a machine learning algorithm effectively mapped the presence of broad-leaf weeds in narrow-leaf crops. For localized management, the model must ideally make as few mistakes as possible. However, when it comes to weed management, it is more acceptable for the model to make errors by indicating the presence of weeds where there is none (FP) rather than the opposite (FN). This more favorable result occurred in Area A, with 4 points classified as FP and only 2 as FN (Table 2). The FP would result in herbicide applications in points where they are not needed. On the other hand, FN would lead to non-application in points that require control, and fewer errors of this type indicate a lower chance of infestation hotspots remaining in the areas, competing for resources with the crop. Thus, in Area A, the model correctly classified 84% of all samples in the test set (accuracy). Of the times the model predicted the presence of weeds, 88% of these predictions were correct (precision). The model correctly predicted 89% of samples with the absence of weeds, i.e., only the presence of pasture (specificity). Additionally, the proportion of true positives (correctly classified positive cases) in relation to the total real positive cases was 79% (recall). The F1 score with a value of 83% indicates a good balance between precision and recall. In the variable importance ranking for the classification model, the first six variables jointly presented 74% importance (Figure 2). The model worsened when using only the six most important variables for classification, increasing the number of errors (Table 3). Therefore, variable selection was dispensed with for this area. Thus, we used all variables to map the presence and absence of weeds in the pasture area (Figure 3).

TABLE 2. Confusion Matrix. Classification of presence-absence of broadleaf weeds in Area A.

		Predicted	
		Presence 1	Absence 0
Observed	Presence 1	17 (TP)	2 (FN)
	Absence 0	4 (FP)	15 (TN)

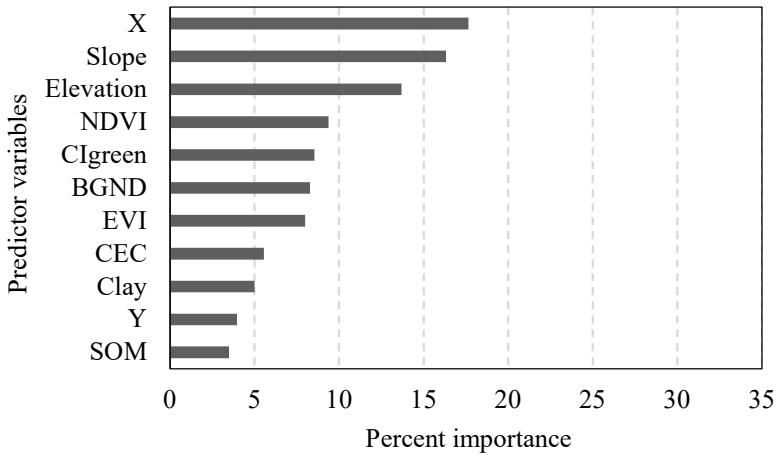

FIGURE 2. Ranking of variable importance for classifying the presence-absence of broadleaf weeds in Area A.

TABLE 3. Confusion Matrix. Classification of the presence-absence of broadleaf weeds in Area A by using the selected variables.

		Predicted	
		Presence 1	Absence 0
Observed	Presence 1	15 (TP)	4 (FN)
	Absence 0	4 (FP)	15 (TN)

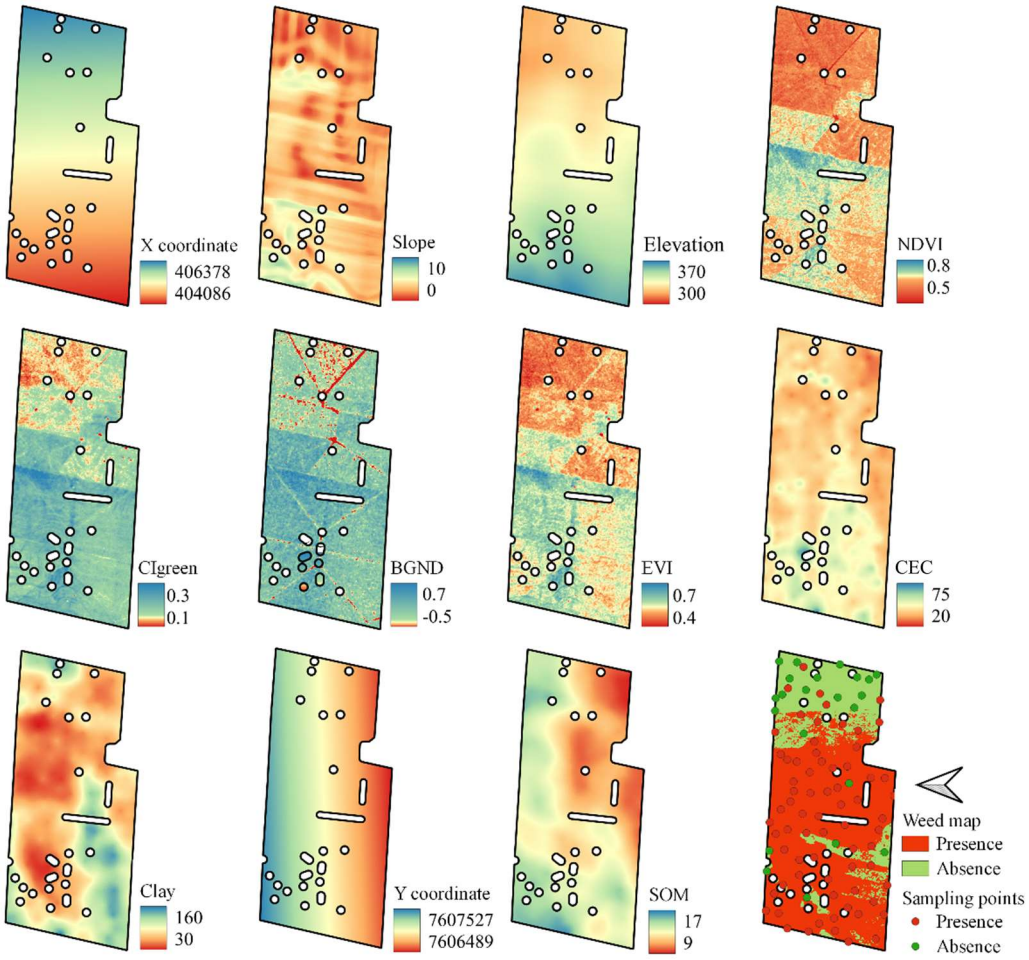

FIGURE 3. Maps of predictor variables and presence-absence map of broadleaf weeds in Area A.

In cultivating sorghum (Area B), the model showed inferior results compared with Area A but still had a satisfactory capacity to classify the presence and absence of weeds (Table 4). The model indicated the absence of weeds where they were present more times, presenting 6 FN (Table 4). Some infested points could potentially miss receiving control. Despite this, it made only one error in indicating presence where they were absent, 1 FP. Regarding the metrics generated from the confusion matrix, the model correctly classified 74% of all samples in the test set (accuracy). Of the times the model predicted the presence of weeds, 67% of these predictions were correct (precision). The model correctly predicted 57% of samples from the absence of weeds class, i.e., only the presence of sorghum (specificity). Additionally, the proportion of true positives (correctly classified positive cases) concerning the total real positive cases was 92% (recall). The F1 score was 77%, indicating a good balance between precision and recall. In the variable importance ranking for the classification model, the first three variables jointly presented 73% importance (Figure 4). The predictive potential of the model remained the same using only three variables as predictors in Area B; therefore, we used them to create the map of the presence and absence of weeds (Figure 5).

TABLE 4. Confusion Matrix. Classification of the presence-absence of broadleaf weeds in Area B.

		Predicted	
		Presence 1	Absence 0
Observed	Presence 1	8 (TP)	6 (FN)
	Absence 0	1 (FP)	12 (TN)

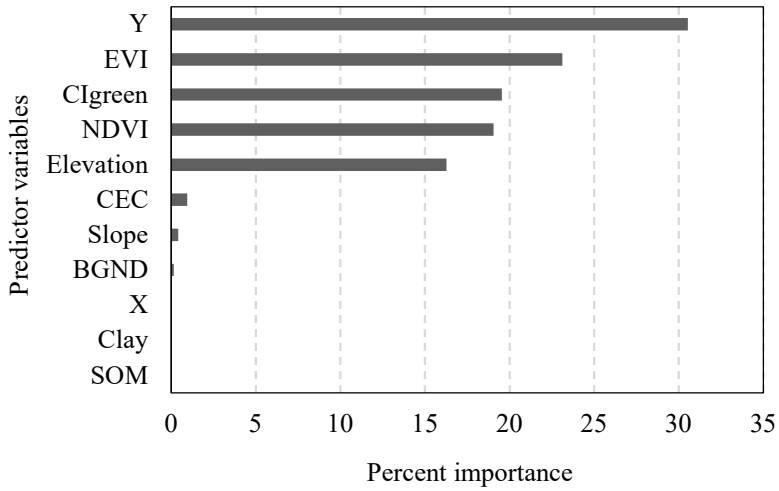

FIGURE 4. Ranking of variable importance for classifying the presence-absence of broadleaf weeds in Area B.

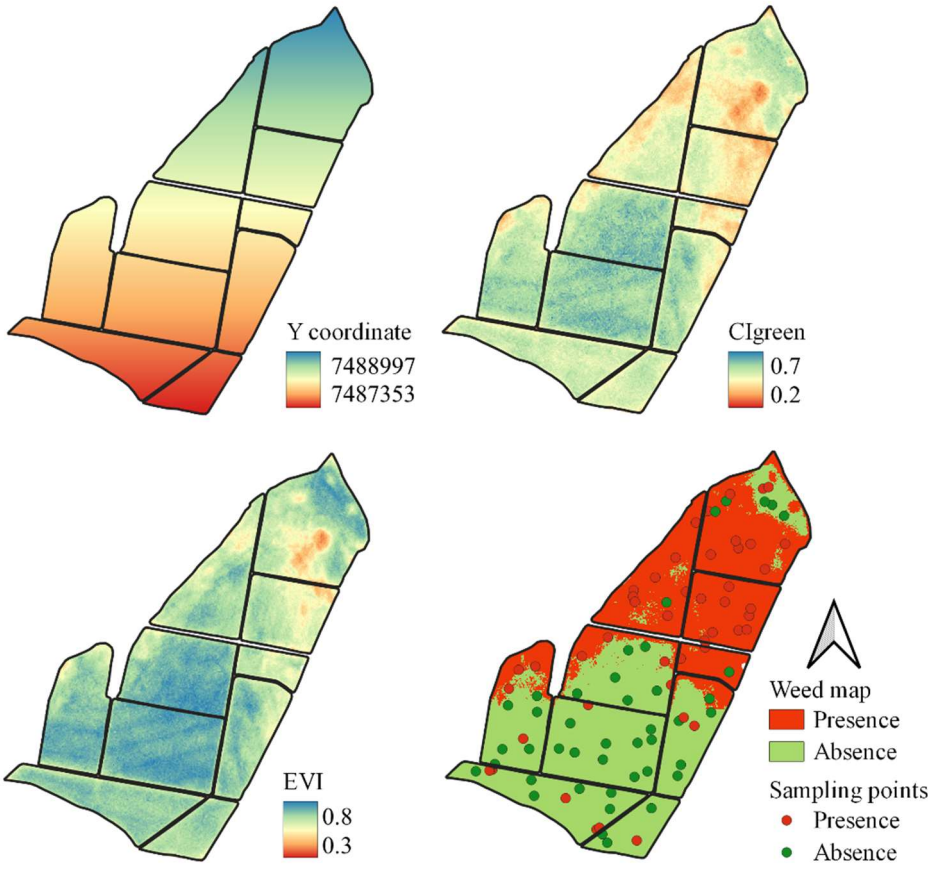

FIGURE 5. Map of selected predictor variables and presence-absence map of broadleaf weeds in Area B.

In both areas, at least one positional variable (X or Y coordinates) was identified as more important (Figures 2 and 4). The significance of these variables in classifications may be associated with specific environmental characteristics, such as wind speed and direction, which influence the formation of weed seed banks in the areas (Pallavicini et al., 2020). Additionally, this may indicate the existence of spatial autocorrelation. In this context, the presence of spatial patterns, such as aggregation commonly observed in the distribution of these plants, favors site-specific weed management (SSWM), where herbicides are applied only in areas where these plants are present (Martín et al., 2015).

Among the vegetation indices, CIgreen, especially in Area B, stood out as one of the most important for classification (Figures 2 and 4). This index is efficient for distinguishing chlorophyll levels among species with distinct characteristics, as is the case with crops such as soybeans (broad leaves) and corn (narrow leaves) (Gitelson et al., 2005). According to these authors, this results from soybeans, besides having a higher amount of chlorophyll on the adaxial surface of the leaves, having a predominantly horizontal leaf arrangement, whereas corn has a more hemispherical shape in the distribution of leaf angle when viewed from above, contributing to such differentiation. Such characteristics are also observed in broad-leaved weeds and crops evaluated in our research. In Area B, the map of the presence and absence of weeds also shows a spatial distribution similar to CIgreen, where points with weeds had lower vegetation index values, i.e., lower biomass (Figure 5). This was indeed observed in the field. In places where weeds were absent, and sorghum was in full development, the crop canopy was closed (Figure 6 A-B), whereas in places infested with weeds, the vegetation was less dense, even showing exposed soil (Figure 6 C-D). This result corroborates with Pallavicini et al. (2020), who observed in cereal crops (narrow leaves) that weeds were predominantly annual smaller in size compared with cultivated plants. In addition, at the time of data collection, the sorghum was more developed than the pasture (Figure 7 A-B). Therefore, in Area B, compared with Area A, this architectural difference observed between the weeds and the sorghum favored their differentiation by spectral data.

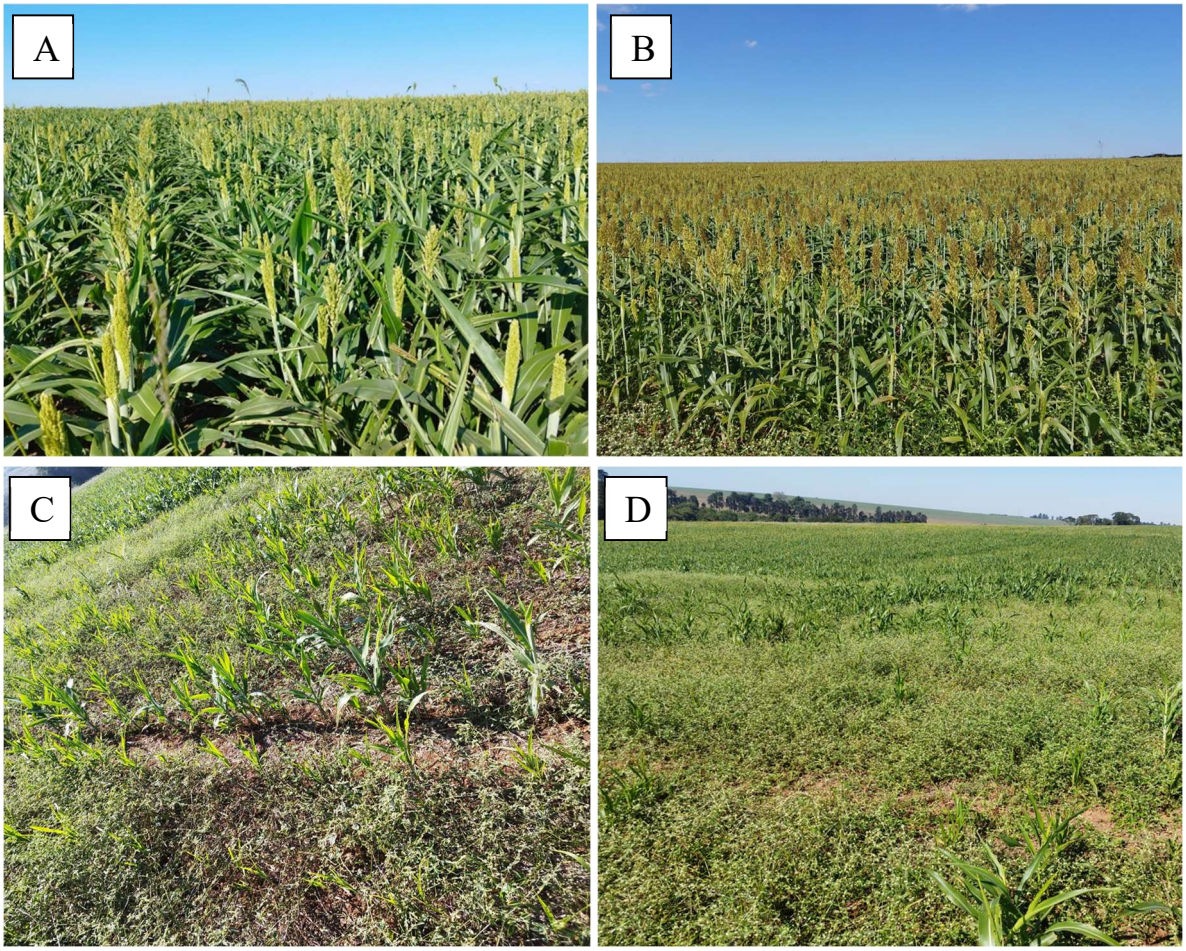

FIGURE 6. Photos of Area B, sorghum crop. A and B – sorghum in full development in locations without broadleaf weeds, and C and D – locations with broadleaf weeds.

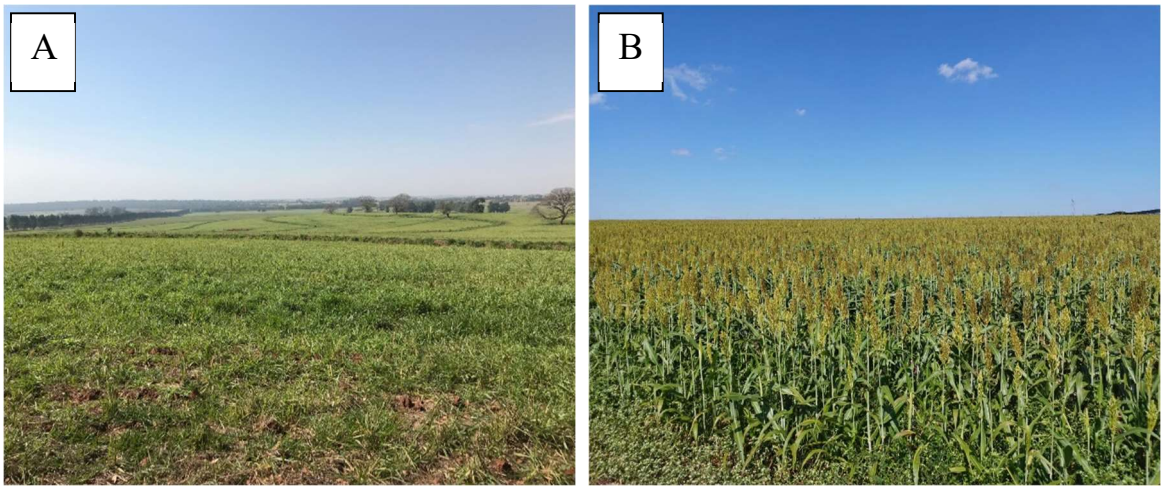

FIGURE 7. Photos of the areas on the sampling dates. A – Area A, pasture. B – Area B, sorghum crop.

Identifying and preventing weed growth in agricultural areas represents a significant challenge, which can increase expenses related to the excessive use of herbicides. For example, Rozenberg et al. (2021) used an RPA to map weeds in onion fields and found that, in five of the 11 analyzed fields, weed coverage was less than 7%. However, according to the authors, herbicide use covered the entire area in all 11 fields. This practice needs to consider weeds' ecology and aggregation characteristics, which rarely cover the entire area. Consequently, excessive spending is incurred even when unnecessary. In our research, weed maps indicated that these plants were present in 70% of the pasture area (Figure 3). On the other hand, in sorghum cultivation, these plants covered only 42% of the area (Figure 5). In this case, especially in sorghum cultivation, applying herbicide only in infested areas could generate savings and avoid excessive use of these pesticides. In practice, employing a digital weed mapping strategy that integrates environmental information as predictors in a machine learning algorithm allows to create maps with reasonably low predictive errors, as observed. Therefore, targeted herbicide applications can be utilized given the specificity of certain active substances for controlling broadleaf weeds (Gutjahr et al., 2012).

However, as observed in our results, we must emphasize that the variables explaining the spatial behavior of weeds vary between cultivation areas. This results from each area presenting unique characteristics, with environmental variables (e.g., slope, wind, texture, moisture, and soil acidity) spatially distributed in diverse ways. Consequently, the influence of these variables on the spatial and temporal patterns of weeds will vary, and environmental variables will influence each weed group differently. In this context, Pätzold et al. (2020) observed that, despite the diversity of weed species varying over ten years of observations, spatial patterns (weed patches) remained stable. This temporal stability suggests that the same environmental variables could be used for classifying weed occurrences in future crops in the same area. However, this will depend on the correct selection of predictor variables in the production area since using variables based on correlations derived from knowledge in other areas can compromise mapping accuracy (Pusch et al., 2023). Therefore, considering that digital data is often readily available, working with a comprehensive set of variables, followed by selection techniques with an evaluation of the metrics generated in predictions, is more advantageous than excluding or selecting variables based solely on the initial perception of the analyst. In this context, when evaluating the metrics of our predictions in an integrated manner, we observed good performance of the models. Thus, the variables we employed proved to be suitable for mapping broadleaf weeds in narrow-leaf crops.

Currently, farmers widely adopt the use of spectral data for weed sampling guidance. However, integrating spectral images and environmental variables as targeted sampling predictors can make these efforts more efficient. Consequently, samples would be collected in locations with a higher probability of weed occurrence, reducing the need to collect numerous samples scattered throughout the area of interest. Furthermore, before crop establishment, pre-emergence application is carried out across the entire area due to the lack of information about where weeds will emerge. With previous studies on the stability of the seed bank of these weeds over time in the production area, the strategy combining variables for weed mapping can also be tested to assist in identifying locations for pre-emergence herbicide application. Combining these techniques with coexistence periods can also benefit integrated weed management, reducing excessive herbicide use and promoting agroecosystem sustainability.

CONCLUSIONS

The combination of machine learning techniques and multiple sources of information related to weed infestations, such as vegetation indices, soil data, topographical characteristics, and spatial information (X and Y coordinates) represents an efficient way to map and manage the occurrence of broadleaf weeds in narrow-leaf crops.

ACKNOWLEDGMENTS

We are very grateful to the owners and managers of Campina and São José Farms for their support and for making their areas available for our research. Also, we would like to thank the members of GITAP for their assistance in data collection. This study was possible due to the São Paulo Research Foundation – FAPESP (Process numbers 2017/50205-9 e 2020/02223-0) financial support and due to the PhD scholarship granted by the National Council for Scientific and Technological Development (CNPq) to the first author.

REFERENCES

Alvares CA, Stape JL, Sentelhas PC, Moraes Gonçalves JL, Sparovek G (2013) Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22(6): 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G, Jones ZM (2016) mlr: Machine Learning in R. Journal of Machine Learning Research 17. https://github.com/mlr-org/mlr

Brenning A, Bangs D, Becker M, Schratz P, Polakowski F (2022) Package 'RSAGA' Type Package Title SAGA Geoprocessing and Terrain Analysis. https://github.com/r-spatial/RSAGA

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16: 321–357. https://doi.org/10.1613/jair.953

Gitelson AA, Viña A, Ciganda V, Rundquist DC, Arkebauer TJ (2005) Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters 32(8): 1–4. https://doi.org/10.1029/2005GL022688

Gutjahr C, Sökefeld M, Gerhards R (2012) Evaluation of two patch spraying systems in winter wheat and maize. Weed Research 52(6): 510–519. https://doi.org/10.1111/j.1365-3180.2012.00943.x

Huete AR, Justice C, Leeuwen W van (1999) MODIS vegetation index algorithm theoretical basis. Environmental Sciences Mod 13. https://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf

Hunter JE, Gannon TW, Richardson RJ, Yelverton FH, Leon RG (2020) Integration of remote-weed mapping and an autonomous spraying unmanned aerial vehicle for site-specific weed management. Pest Management Science 76(4): 1386–1392. https://doi.org/10.1002/ps.5651

Lamb DW, Brown RB (2001) Remote-sensing and mapping of weeds in crops. Journal of Agricultural and Engineering Research 78(2): 117–125. https://doi.org/10.1006/jaer.2000.0630

Martín CS, Andújar D, Fernández-Quintanilla C, Dorado J (2015) Spatial distribution patterns of weed communities in corn fields of central Spain. Weed Science 63(4): 936–945. https://doi.org/10.1614/ws-d-15-00031.1

Metcalfe H, Milne AE, Coleman K, Murdoch AJ, Storkey J (2019) Modelling the effect of spatially variable soil properties on the distribution of weeds. Ecological Modelling 396: 1–11. https://doi.org/10.1016/j.ecolmodel.2018.11.002

Nordmeyer H (2006) Patchy weed distribution and site-specific weed control in winter cereals. Precision Agriculture 7(3): 219–231. https://doi.org/10.1007/s11119-006-9015-8

Pallavicini Y, Plaza EH, Bastida F, Izquierdo J, Gallart M, Gonzalez-Andujar JL (2020) Weed seed bank diversity in dryland cereal fields: does it differ along the field and between fields with different landscape structure? Agronomy 10(4): 1–14. https://doi.org/10.3390/agronomy10040575

Pätzold S, Hbirkou C, Dicke D, Gerhards R, Welp G (2020) Linking weed patterns with soil properties: a long-term case study. Precision Agriculture 21(3): 569–588. https://doi.org/10.1007/s11119-019-09682-6

Pusch M, Samuel-Rosa A, Magalhães PSG, Amaral LR (2023) Covariates in sample planning optimization for digital soil fertility mapping in agricultural areas. Geoderma 429: 116252. https://doi.org/10.1016/j.geoderma.2022.116252

Rouse JW, Haas RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the great plains with ERTS. Washington, DC, USA, p.309-317. Available: https://ntrs.nasa.gov/citations/19740022614

Rozenberg G, Kent R, Blank L (2021) Consumer-grade UAV utilized for detecting and analyzing late-season weed spatial distribution patterns in commercial onion fields. Precision Agriculture 22(4): 1317–1332. https://doi.org/10.1007/s11119-021-09786-y

Sa I, Popović M, Khanna R, Chen Z, Lottes P, Liebisch F, Nieto J, Stachniss C, Walter A, Siegwart R (2018) WeedMap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sensing 10(9). https://doi.org/10.3390/rs10091423

Souza MF de, Amaral LR do, Oliveira SR de M, Coutinho, MAN, Ferreira Netto C (2020) Spectral differentiation of sugarcane from weeds. Biosystems Engineering 190: 41-46. https://doi.org/10.1016/j.biosystemseng.2019.11.023

Szatmári G, László P, Takács K, Szabó J, Bakacsi Z, Koós S, Pásztor L (2019) Optimization of second-phase sampling for multivariate soil mapping purposes: case study from a wine region, Hungary. Geoderma 352: 373–384. https://doi.org/10.1016/j.geoderma.2018.02.030

Thorp KR, Tian LF (2004) A review on remote sensing of weeds in agriculture. Precision Agriculture 5(5): 477–508. https://doi.org/10.1007/s11119-004-5321-1

Wadoux AMJC, Minasny B, McBratney AB (2020) Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Science Reviews 210: 103359. https://doi.org/10.1016/j.earscirev.2020.103359