Pre-planting weed detection based on ground field spectral data

Luan Pierre Pott [a,d,†], Telmo Jorge Carneiro Amado [b], Raí Augusto Schwalbert [a,d], Elodio Sebem [c], Mithila Jugulam [d], Ignacio Antonio Ciampitti [d, ‡]

[a] Agricultural Engineering Department, Federal University of Santa Maria, Rural Science Centre, Santa Maria, RS, Brazil;

[b] Soil Department, Federal University of Santa Maria, Rural Science Centre, Santa Maria, RS, Brazil;

[c] Technology in Geoprocessing, Federal University of Santa Maria, Polytechnic School, Santa Maria, RS, Brazil;

[d] Department of Agronomy, Kansas State University, 2004 Throckmorton Plant Science Center, Manhattan, KS 66506, USA;

Running title: Pre-planting spectral weed detection

E-mail address:

Luan P. Pott: luanpierrepott@hotmail.com

Telmo J. C. Amado: proftelmoamado@gmail.com

Raí A. Schwalbert: rai.schwalbert@gmail.com

Elodio Sebem: elodiosebem@politecnico.ufsm.br

Mithila Jugulam: mithila@ksu.edu

Ignacio A. Ciampitti: ciampitti@ksu.edu

[†] Corresponding author. Luan P. Pott; luanpierrepott@hotmail.com

[‡] Corresponding author. Ignacio A. Ciampitti; ciampitti@ksu.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ps.5630

Abstract:

BACKGROUND: Site-specific weed management (SSWM) demands higher resolution data for mapping weeds in fields, but the success of this tool relies on the efficiency of optical sensors to discriminate weeds relative to other targets (soils, and residues) before cash crop establishment. The objectives of this study were to i) evaluate the accuracy of spectral bands to differentiate weeds (target) and other non-targets; ii) access vegetation indices (VIs) to assist in the discrimination process; and iii) evaluate the accuracy of the thresholds to distinguish weeds relative to non-targets for each VI using training and validation data sets.

RESULTS: The main outcomes of this study for effectively distinguishing weeds from other non-targets are: (i) training and validation data exhibited similar spectral curves; (ii) red and near-infrared (NIR) spectral bands presented greater accuracy relative to the other bands; (iii) the tested VIs increased the discrimination accuracy related to single bands, with an overall accuracy above 95% and a kappa above 0.93.

CONCLUSION: This study provided a novel approach to distinguish weeds from other non-targets utilizing a ground-level sensor before cash crop planting based on field spectral data. However, the limitations of this study are related to the spatial resolution to distinguish weeds that might be closer to the one this study presented, and also related to the soil and crop residues conditions at the time of collecting the readings. Overall the results presented contribute to an improved understanding of spectral signatures from different targets (weeds, soils, and residues) before planting time supporting SSWM.

Keywords: site-specific weed management (SSWM); spectral curves, spectral bands; vegetation indices.

1 Introduction

Weeds reduce crop growth and yield by competing with field crops for environmental resources such as light, water, and nutrients.[1,2,3] The cash crop yield reduction is increased when weed competition occurs in early growth stages. In overall, weed infestation contribute to the largest potential yield losses (34%) relative to other plant pests such as insects or pathogens.[4,5]

Generally, the distribution of weeds in a crop field are in patches;[6-7] however, herbicides are applied uniformly across fields. Chemical methods such as the use of herbicides is the most effective for weed control in modern agriculture.[9] The concept of site-specific weed management (SSWM) has the bias to cope with the heterogeneity occurring within fields by treating only weed patches.[10] The SSWM may result in lower use of herbicides with reduction in environmental hazard, and can save input costs, depending on the level of weed infestation.[10] Previous studies reported that with SSWM, herbicide savings range from 53 to 75% relative to uniformly applied herbicide across the field.[11,12]

Conventional weed mapping operations are costly in commercial large-scale farms and time-consuming in small farms.[13,14] Use of remote sensing, more specifically, sensor-based systems can offer an alternative (e.g., SSWM) to conventional weed mapping. Use of the ground-level hyperspectral systems to scout weeds presents benefits such as non-contact detection, fast response, high reliability, and low power consumption, making this method a simple and easy real-time procedure.[15-19]

Distinguishing weeds relative to other non-targets, such as crop residues and types of soil (e.g., textural classes), for pre-planting applications is of great interest to improve weed

identification and therefore, effective weed control in site specific management. Previous studies have found that the use of spectral bands help to differentiate plants from other non-targets.[20] Huete *et al*. reported the spectral response of a plant canopy with different soil backgrounds.[21] Langner *et al*. utilizing a camera, reported a special vegetation index (VI), which has a red threshold criterion aggregate in the normalized difference vegetation index (NDVI) calculation.[22] Scotford and Miller investigated the utilization of spectral reflectance techniques using vegetation and soil spectral curves.[16] Therefore, testing spectral bands and/or VIs and setting thresholds for distinguishing weeds using ground-level sensors is a key strategy to improve identification of weedy patches.[15]

Plant leaves and canopy, in general, are mainly affected by plant pigments including chlorophyll (e.g., chlorophyll a and b), carotenes and xanthophylls in visible light reflectance.[23-25] The combination of chlorophyll, a strong scattering of light, and internal cellular plant structure affect the red-edge band reflectance.[24,26] In the near infrared (NIR) band, internal leaf structure and multiple leaf layers influence the reflectance properties of the canopy.[23,24,27] In senescent plants, the chlorophyll can gradually show a decrease in the content which has greater reflectance in blue and red bands.[25,28,29]

Recording hyperspectral information about different targets when scouting for weeds can be a challenge, due to the high amount of auto correlated data. For that reason, in recent years, multivariate analyses, such as linear discrimination, have been employed to discriminate crops versus weeds.[30-32] Decision tree-based analyses are another class of algorithms that have the potential for helping researches to interpret hyperspectral data. Conditional inference tree is a decision tree algorithm for recursive binary splitting. It embeds the framework in a well-defined statistical environment based on permutation tests,

attempting to distinguish between significant and insignificant improvements.[33] Random forests algorithm is one of the most powerful machine-learning techniques.[34] Recent studies utilized random forest in remote sensing with multiple applications such as land cover classification, tree species mapping, and vegetation classification.[27,35,36] Therefore, these techniques can greatly contribute to the identification of spectral bands for discriminating weeds versus other non-targets such as soil and plant residue.

The objectives of this study were to: i) evaluate the accuracy of spectral bands to discriminate weeds and other targets (soils and residues) in pre-planting applications utilizing ground-level sensing; ii) access VIs to assist in the discrimination between weeds and other non-targets; iii) evaluate the accuracy improvement of the thresholds to distinguish the weeds relative to other non-targets for each VI using training and validation data sets.

2. Materials and Methods

2.1 Training dataset

Two sources of data were collected and used in this study, i) field trials (herein termed as 'training data'), and ii) on-farm (herein termed as 'validation data') data. Training data were collected from a field trial carried out during the 2016/17 soybean growing season (November 2016 to April 2017) at the Federal University of Santa Maria, Rio Grande do Sul, Brazil. The study comprises of eight species (seven weeds and one cash crop): *Amaranthus hybridus* L., *Bidens pilosa* L., *Brachiaria plantaginea* (Link) Hitchc., *Euphorbia heterophylla* L., *Glycine max* (L.) Merr. (soybean), *Ipomoea grandifolia* (Dammer) O'Donell, *Panicum maximum* Jacq., and *Sida rhombifolia L.* The experimental

design was completely randomized with four replications. The plot dimensions were 3 x 2 m. All species were sown on November 10th in a four-row arrangement with 0.5 m between rows. The plant density was adjusted to 300,000 plants ha^{-1} for each specie trough manual plant thinning.

In order to perform the crop residue (mulch) readings, corn (*Zea mays* L.), soybean, and black oat (*Avena sativa* L.) were all grown during the 2016/17 season, and after crop harvest, the measurements were done. Measurements in all three types of crop residues were collected to form the target residue. In the field trials, we collected soil samples at 0.10 m depth to identify the soil texture, and classified as sandy soil (clay 18%, silt 22% and sand 60%), according to the United States Department of Agriculture (USDA) soil texture triangle.[37]

2.2 Validation dataset

The validation data set was built by collecting radiometer readings at five on-farm sites located in a radius of 200 km (Não-Me-Toque, Tapera, Cruz Alta, Júlio de Castilhos and Itaara counties) (Fig. 1). Evaluations were carried in those fields from February 2016 until March 2018. The radiometer readings of weeds were collected over two years, in both summer and winter crops, to compose the data set for validation. In addition, radiometer readings were collected from sandy soil sample, residues as well as a different soil type (clay 58%, silt 22%, and sand 20%), classified as clay soil, based on the soil texture triangle.[37]

2.3 Data collection

Reflectance was measured utilizing the FieldSpec® HandHeld 2™ ASD Inc. spectroradiometer (passive sensor with evaluations collected in the zenith position, 0.3 m from the target) ranging from 325 to 1075 nm, with an accuracy of ±1 nm and a resolution of <3 nm at 700 nm. The sensor used to collect all the spectral data has a 25-degree full conical angle field-of-view, capable of collecting spectral data of 277.71 cm² for each measure. The readings were performed in points only on sunny days, in the absence of clouds at noon (12 pm). Before starting the measurements, and every half hour the sensor was properly calibrated with Spectralon® board following the manufacturer recommendations.[38]

In the field trials (training data), all weeds and crops were evaluated at three different phenological growth stages according to the BBCH scale:[39] stage 13 (3 leaves); stage 17 (7 leaves); and stage 51 (inflorescence or flower buds visible). For all the on-farm sites, plants phenological growth stages ranged from 11 (1 leaf) to 59 (first flower petals visible).

For the field research data, 393 radiometer weed readings were collected, and for the on-farm fields, 248 weed readings. Additionally, 75 radiometer readings were collected from crop residues (soybean, corn, and black oats mulch), 30 readings from clay soil, and 39 readings from sandy soil in field research and on-farm fields (Table 1). The readings were collected taking full (100%) of field-of-view of each target/non-target.

2.4 Data analyses

The framework of the data analyses is presented in Figure 2. The Linear discriminant analysis (LDA) (only using training data) was executed to separate classes among the targets evaluated in this study when using the full electromagnetic spectrum. Discriminant analysis,

a multivariate technique, was utilized to separate groups based on the measured k variables in each sample, finding one or more linear combinations of the selected variables.[40] The *ade4* package was used to perform the analysis within the R environment.[41,42]

Considering that most of the spectral sensors developed for agricultural purposes (e.g., GreenSeeker®, WeedSeeker®, OptRx®, WEEDit®) have a limited radiometer resolution, recording only a few bands or VIs relative to the FieldSpec, individual wavelengths (recorded by FieldSpec) were grouped in order to form spectral bands (Table 2). The definition of the spectral bands was based on the spectral bands from WorldView-2, which agrees with the critical spectral bands identified for future spectral capabilities to satellites.[43]

All spectral bands (only using training data) were subjected to conditional inference tree analysis within the *partykit* package in R in order to select spectral bands that allow a more evident separation between weeds and other targets.[44] These analysis are based on hierarchically ordered and recursively repeated binary splits, where the strength of each association is measured by a P-value. The terminal nodes account for the final subset of density points in each target. One of the main advantages of this technique is the possibility of exploration of complex interactions with control over overfitting issues.[45]

Random forest analysis (using both training and validation data) was performed using *randomForest* package in R to analyze the separation capacity of the targets by the selected bands.[46] Mean decrease accuracy (MDA) was assessed to select the most important spectral bands to accomplish VIs. The MDA utilizes permuting out-of-bag (OOB) samples to compute the importance of the variable. The OOB sample is the set of observations which are not used for building the current tree. It is used to estimate the prediction error and to

evaluate variable importance.[46,47] The OOB is the mean prediction error on each training sample, using only the trees that did not have in their bootstrap sample.[47]

To test if the data collected in the field trial is comparable with the one gathered from the on-farm sites, the average, minimum and maximum spectral curves were plotted.

Lastly, the VIs: enhanced vegetation index (EVI), enhanced vegetation index 2 (EVI2), optimized soil adjusted vegetation index (OSAVI), soil adjusted vegetation index (SAVI), normalized difference vegetation index (NDVI), difference between NIR and RED (NIR-RED), and ratio between NIR and RED (NIR/RED) (Table 3) were determined to test the potential of combination of the different spectral bands as well as to improve the accuracy in weed discrimination relative to the other targets, once no commercial sensors may calculate and compare all those VIs. As a criterion, overall accuracy and kappa coefficients were performed using the *caret* package.[48]

3 Results

3.1 Spectral curves

Reflectance readings (average) and its variation (minimum to maximum) for each target, i.e., weeds, residue, clay, and sandy soils were recorded (Fig. 3). The living plants (weeds) presented the highest reflectance in the green band (501-565 nm) and absorption peaks at blue (441-485 nm) and red (626-690 nm) bands in the visible light. The red edge (691-750 nm) band presented a broad rate of variation (minimum to maximum) in the spectral curves. The NIR band presented the highest reflectance for weeds, with a broad range of variation, relative to the soil texture (clay and sandy) and residue types (Fig. 3). The crops residues and soil types presented the narrowest variation on the spectral

reflectance curves, with the reflectance rising as the wavelength increases from 400 nm onwards (to 900 nm). Among the soils type, sandy soil texture presented the highest reflectance values compared to clay soil.

3.2 Comparison between field trial and on-farm fields

The reflectance for weeds, crop residues, and sandy soil were greater for the field trial relative to the on-farm sites, but overall there was an overlap between the curves without reflecting a significant difference in mean values of the spectral curves for the classes between these two data sources evaluated in this study (Fig. 4). The target clay soil was collected only in on-farm fields.

3.3 Linear discriminant analysis (LDA)

The LDA was able to differentiate all the targets (weeds, residue, and sandy soil) (Fig. 5). There was no overlapping among classes, indicating a high degree of separation when considering the full-electromagnetic spectrum.

3.4 Conditional inference tree analysis

Conditional inference tree analysis resulted in four inner nodes (1, 2, 5, and 7) and five terminal nodes (3, 4, 6, 8, and 9) (Fig. 6). The spectral bands differentiating classes were: red, NIR, red edge and green (Fig. 6). The class for weeds was identified and separated by the bands red (inner node 1) and NIR (inner node 2). The conditions applied to classify the target as weeds were: red ≤ 0.13 and NIR > 0.233. Following these conditions (nodes), a total of 99.5% of all the radiometric readings from the weeds were obtained in the terminal

node 4. For the sandy soil, three spectral bands were used: red (inner node 1), red edge (inner node 5) and green (inner node 7). For this soil texture, the classification conditions were red > 0.13, red edge > 0.266 and green ≤ 0.144, obtaining 100% of the total sandy soil database in the terminal node 8. Residues were obtained overlapping of targets in their classification (nodes 3, 6 and 9). For most of the crop residue (77.3%), the classification occurred with the bands: red > 0.13 and red edge ≤ 0.266, terminal node 6 (Fig. 6).

3.5 Random forest analysis (training and validation data studies)

For the random forest analysis confirmed for both data sets that the red and NIR bands are of great relevancy when identifying weeds versus other targets, as reported by the MDA (Fig. 7). After removing the red spectral band, the model accuracy decreased to the value 31.0% followed by the spectral band NIR with the accuracy 19.8% (Fig 7). The other spectral bands presented lower impacts on model performance relative to both red and NIR bands.

3.6 Vegetation indices

The VIs: EVI, EVI2, OSAVI, SAVI, NDVI, NIR-RED, and NIR/RED improved the discrimination of weeds relative to the other targets in comparison with the utilization of the isolated spectral bands. The values of each index had no overlap between weeds and the other targets (Fig. 8) for the training data.

Conditional inference trees with training data were developed for each index separately in order to compare their capacity of discriminating weeds from other targets and to establish VI-specific thresholds (Fig. 9). The EVI had 0.138 as a threshold value, EVI2 had 0.221,

OSAVI had 0.166, SAVI had 0.162, NDVI had 0.197, NIR - RED had 0.137, and NIR/RED had 1.485.

The thresholds from the training data were applied over the validation data in order to calculate the overall accuracy and kappa coefficients from the confusion matrix. The OSAVI had 100% of overall accuracy and kappa equal 1, and the other indices: EVI, EVI2, SAVI, NDVI, NIR – RED and NIR/RED also presented high values of accuracy, with overall accuracy > 95%, and kappa > 0.93 (Table 4).

4 Discussion

Remote sensing of weeds relative to other targets before soybean planting is a newer approach with potential of weed management improvement by reducing environmental risk and cutting cost. To full perform SSWM, it is necessary to discriminate accurately weeds from other targets in order to allow only target spraying (applying herbicide only where it is necessary).[8] Discrimination of weeds relative to other targets can be assessed at pre-planting and, in early and late during the growing season, at post-emergence applications. Researchers have been studying alternatives to discriminate plants relative to soil targets, such as the use of different VIs from multispectral cameras.[22,51,52] More recently, remote sensing was used to distinguish weeds relative to cash crops, such as corn, cabbage (*Brassica olearacea* var. *capitata*), sugar beet (*Beta vulgaris* L.) and soybean.[15,18,27,32,55,56] Several studies have provided insights for developing sprayers based on the SSWM approach, with limitations mainly related to identify weeds and the spatial resolution for discriminating other non-target.[6,72,73]

Considering the scientific literature on spectral information for different targets before planting, an important outcome of this study was related to the similarity of spectral curves between training (field trial) and validation (on-farm) data sets, to assess weeds, soil texture, and crop residues. This relevant finding shows that hyperspectral data collected in experimental and on-farm conditions (training and validation data) can be used to develop models to discriminate weeds from other targets in field condition. Plant canopy reflectance is governed by the concentration and distribution of biochemical compounds, internal structure of the tissue, as well as the leaf surface properties.[27] Daughtry reported that spectral curves of three different crop residues and five soil types had similar behavior but differ only in their amplitude of reflectance.[57] The reflectance of soils and crop residues rising as the wavelength increase from 400 nm onwards to 900 nm. Thus, the reflectance is generally lower in the visible range (400-690 nm) and higher in the NIR region (691-900 nm) in these selected targets.[58] Soil texture has high reflectance for sandy soils, most likely due to the high amount of quartz in the sand fraction, increasing the intensity of spectral reflectance.[58,59]

The second outcome of this study was connected to perfect discrimination among weeds and the other non-targets - residues, sandy and clay texture soils - using all wavelengths reported by the LDA approach. Spectral curves often carry a more detailed information compared to wavelength aggregated in spectral bands.[27] The LDA has been recently used for hyperspectral data target classification.[30-32] Conditional inference tree and random forest analyses reported that red and NIR bands were the most important spectral bands to discriminate weeds relative to other targets. Similar results were reported by Rondeaux *et al*. in a study with vegetation and soil reflectances.[51] According to Langner *et*

al. the reflectance of green vegetation is very low in the red band, in contrast with the high reflectance in NIR.[22] Soils on the other hand, have a reflectance with minimum difference between red and NIR bands. The results are in line with the commercial sensors that use the red and NIR spectral bands to weed identification.[60]

The selection of VIs was defined related to their relevancy for agricultural purposes, composing the spectral bands which had the greater values to distinguish weeds relative to other targets (spectral bands: red and NIR).[61] Using VIs relative to the individual spectral bands improved the performance (overall accuracy and kappa coefficient) for discriminating weeds. The third outcome of this study was related to the ability of the indices: EVI, EVI2, OSAVI, SAVI, NDVI, NIR - RED and NIR/RED to discriminate weeds from other targets. The VIs are related to several properties of plants and those VIs are frequently used to other aims as disease detection, plant stress, nutrition, yield forecast and phenology.[61-66]

From all the VIs tested on this study, OSAVI presented the greatest accuracy when comparing training and validation data sets. The OSAVI was built to optimized soil-adjusted vegetation index (SAVI) with the aim of reducing the sensitivity of the NDVI to soil background and atmospheric effects.[51,52] Recently, the OSAVI has also been used for canopy stress detection, chlorophyll content estimation, monitoring nitrogen (N) status and estimating vegetal biomass and canopy coverage.[67-71] However, all the VIs tested on this study presented high values of accuracy, effectively distinguishing weeds relative to other targets before planting (with 277.71 cm^2 of spatial resolution).

One of the constraints of this study was related to the spatial resolution utilized with the sensor for collecting the spectral data, limiting the extrapolation of this information to lower scale of spatial resolution. Another limitation is related to the soil and crop residues

conditions at the time of collecting the readings. The spectral curves of soil and residues were determined by the cumulative property resulting from heterogenic combinations of mineral and organic material and their moisture.[74,75] Daughtry showed that crop residues and soil have similar behavior for the spectral curves under different water content, but both targets presented an increase in the reflectance values when the presented contrasting moisture levels (dry vs. wet conditions).[57] This complex interaction between the targets and the environmental conditions should be explored in future studies.

The next step on SSWM studies will be to aggregate different spatial resolutions with multiples platforms (such as satellites, unmanned aerial vehicle (UAV) and proximal sensors) to analyze the influence the VIs when considered several weed densities in pre-planting cash crop. In addition, future investigations should look into the performance of the remote sensing tools for identifying weeds when considering more environmental and soil factors as variables to be included in the analyses.

5 Conclusions

The results of this research provided useful pre-planting data source for distinguishing weeds related to other targets using high spatial resolution ground-sensing. The findings could support further field applications of spectral data in the field for improving SSWM.

The main outcomes of this study were that: (i) spectral curves of weeds, sandy soil, and crop residues were similar for both training and validation data sets permitting evaluate a large database for discriminating weeds relative to other targets; (ii) spectral bands, red and NIR had greater values of accuracy to discriminate weeds relative to crop residues, sandy and clay texture soils; (iii) the VIs: EVI, EVI2, OSAVI, SAVI, NDVI, NIR - RED and

NIR/RED had the greater values of accuracy to discriminate weeds relative to others targets when compared with single spectral bands.

The thresholds of VIs defined in this study might provide values of classification of weeds relative to other targets in the fields using multiples alternatives sensors according with the presented results.

Author Contributions: Conceptualization, L. P. P., T. J. C. A. and I. A. C.; L. P. P. led the writing of the paper; L. P. P. and R. A. S. lead the statistical analysis; T. J. C. A. and R. A. S. contributed to the data discussion; E. S., M. J. and I. A. C. contributed in writing-review and editing the paper; I. A. C. led the study and contributed to the data analysis/discussion.

Funding:

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES) Finance Code 001. Aquarius project (http://w3.ufsm.br/projetoaquarius/index.php/pt/) and Kansas Corn Commission.

Acknowledgments:

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES) Finance Code 001. Aquarius project (http://w3.ufsm.br/projetoaquarius/index.php/pt/) and Kansas Corn Commission. This is contribution no. 19-235-J from the Kansas Agricultural Experiment Station. The authors thank the farmers for their cooperation with this study.

Conflicts of Interest:

The authors declare no conflict of interest.

References

1. Harper, J. L. (1977). The population biology of plants. Academic Press, London, UK.

2. Swanton CJ, Nkoa R, and Blackshaw RE, Experimental Methods for Crop–Weed Competition Studies, *Weed Sci* **63**:2–11 (2015).

3. Jha P, Kumar V, Godara RK, and Chauhan BS, Weed management using crop competition in the United States: A review, *Crop Prot* **95**:31–37, Elsevier Ltd (2017).

4. Oerke EC, Crop losses to pests, *J Agric Sci* **144**:31–43 (2006).

5. Gharde Y, Singh PK, Dubey RP, and Gupta PK, Assessment of yield and economic losses in agriculture due to weeds in India, *Crop Prot* **107**:12–18, Elsevier (2018).

6. Slaughter DC, Giles DK, and Downey D, Autonomous robotic weed control systems: A review, *Comput Electron Agric* **61**:63–78 (2008).

7. Pantazi XE, Moshou D, and Bravo C, Active learning system for weed species recognition based on hyperspectral sensing, *Biosyst Eng* **146**:193–202, Elsevier Ltd (2016).

8. Shirzadifar A, Bajwa S, Mireei SA, Howatt K, and Nowatzki J, Weed species discrimination based on SIMCA analysis of plant canopy spectral data, *Biosyst Eng* **171**:143–154, Elsevier Ltd (2018).

9. Harker KN and O'Donovan JT, Recent Weed Control, Weed Management, and Integrated Weed Management, *Weed Technol* **27**:1–11 (2013).

10. Andújar D, Rueda-Ayala V, Moreno H, Rosell-Polo JR, Escolà A, Valero C, *et al.*,

Discriminating crop, weeds and soil surface with a terrestrial LIDAR sensor, *Sensors (Switzerland)* **13**:14662–14675 (2013).

11. Gerhards R and Christensen S, Real-time weed detection, decision making and patch spraying in maize, sugarbeet, winter wheat and winter barley, *Weed Res* **43**:385– 392 (2003).

12. Young DL, Kwon TJ, Smith EG, and Young FL, Site-specific herbicide decision model to maximize profit in winter wheat Special Issue Number 3 on the European Conferences on Precision Agriculture, *Precis Agric* **4**:227–238 (2003).

13. Rew LJ and Cousens RD, Spatial distribution of weeds in arable crops: Are current sampling and analytical methods appropriate?, *Weed Res* **41**:1–18 (2001).

14. Barroso J, Fernandez-Quintanilla C, Maxwell BD, and Rew LJ, Simulating the effects of weed spatial pattern and resolution of mapping and spraying on economics of site-specific management, *Weed Res* **44**:460–468 (2004).

15. Peteinatos GG, Weis M, Andújar D, Ayala VR, Gerhards R, Potential use of ground-based sensor technologies for weed detection, *Pest Manag Sci* **70**:190-199 (2013).

16. Scotford IM and Miller PCH, Applications of spectral reflectance techniques in northern European cereal production: A review, *Biosyst Eng* **90**:235–250 (2005).

17. Huang Y, Lee MA, Thomson SJ, and Reddy KN, Ground-based hyperspectral remote sensing for weed management in crop production, *Int J Agric Biol Eng* **9**:98–109 (2016).

18. Akbarzadeh S, Paap A, Ahderom S, Apopei B, and Alameh K, Plant discrimination by Support Vector Machine classifier based on spectral reflectance, *Comput Electron Agric* **148**:250–258 (2018).

19. Rogalski A, Infrared detectors: Status and trends, *Prog Quantum Electron* **27**:59–210 (2003).

20. Wang N, Zhang N, Peterson DE, and Dowell FE, Design of an optical weed sensor using plant spectral characteristics, *Biolog. Qual. Prec. Agric. II* ***4203***: 409-419 (2001).

21. Huete AR, Jackson RD, and Post, DF, Spectral response of a plant canopy with different soil backgrounds, *Remote Sens Environ* **17:**37–53 (1985).

22. Langner HR, Böttger H, and Schmidt H, A special vegetation index for the weed detection in sensor based precision agriculture, *Environ Monit Assess* **117**:505–518 (2006).

23. Gausman H. Plant leaf optical properties. Texas Tech Press, Lubbock, Texas. (1985).

24. Fletcher RS and Reddy KN, Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds, *Comput Electron Agric* **128**:199–206 (2016).

25. Dordas C, Nitrogen nutrition index and leaf chlorophyll concentration and its relationship with nitrogen use efficiency in barley (Hordeum vulgare L.), *J Plant Nutr* **40**:1190–1203 (2017).

26. Ray T, Murray B, Chehbouni A, and Njoku E, The red edge in arid region vegetation: 340–1060 nm spectra. Summaries of the 4th Annual JPL Airborne Geoscience Workshop. Ed.1. AVIRIS Workshop. Jet Propulsion Laboratory, Pasadena, CA (1993).

27. Gao J, Nuyttens D, Lootens P, He Y, and Pieters JG, Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery, *Biosyst Eng* **170**:39–50 (2018).

28. Dordas CA and Sioulas C, Safflower yield, chlorophyll content, photosynthesis, and water use efficiency response to nitrogen fertilization under rainfed conditions, *Ind Crops Prod* **27**:75–85 (2008).

29. Ziadi N, Brassard M, Bélanger G, Claessens A, Tremblay N, Cambouris AN, *et al.*, Chlorophyll measurements and nitrogen nutrition index for the evaluation of corn nitrogen status, *Agron J* **100**:1264–1273 (2008).

30. Imani M and Ghassemian H, Two Dimensional Linear Discriminant Analyses for Hyperspectral Data, *Photogramm Eng Remote Sens* **81**:777–786 (2015).

31. Henry WB, Shaw DR, Reddy KR, Bruce LM, and Tamhankar HD, Spectral reflectance curves to distinguish soybean from common cocklebur (Xanthium strumarium) and sicklepod (Cassia obtusifolia) grown with varying soil moisture, *Weed Sci* **52**:788-796 (2004).

32. Gray CJ, Shaw DR, and Bruce LM, Utility of Hyperspectral Reflectance for Differentiating Soybean (Glycine max) and Six Weed Species, *Weed Technol* **23**:108–119 (2009).

33. Sardá-Espinosa A, Subbiah S, and Bartz-Beielstein T, Conditional inference trees for knowledge extraction from motor health condition data, *Eng Appl Artif Intell* **62**:26–37 (2017).

34. Breiman, L, Random forests, *Mach Learn* **45**:5-32 (2001).

35. Belgiu M and Drăgu L, Random forest in remote sensing: A review of applications and future directions, *ISPRS J Photogramm Remote Sens* **114**:24–31 (2016).

36. Nitze I, Barrett B, and Cawkwell F, Temporal optimisation of image acquisition for land cover classification with random forest and MODIS time-series, *Int J Appl Earth*

Obs Geoinf **34**:136–146, Elsevier B.V. (2015).

37. Shirazi MA and Boersma L, A Unifying Quantitative Analysis of Soil Texture1, *Soil Sci Soc Am J* **48**:142 (2010).

38. FieldSpec® HandHeld 2™ Spectroradiometer User Manual. ASD Document 600860 Rev. D. (2010).Available online: http://www.geo-informatie.nl/courses/grs60312/material2017/manuals/600860-dHH2Manual.pdf accessed on 02/02/2019)

39. Hess M, Barralis G, Bleiholder H, Buhr L, Eggers T, Hack H, *et al.*, Use of the extended BBCH scale - General for the descriptions of the growth stages of mono- and dicotyledonous weed species, *Weed Res* **37**:433–441 (1997).

40. Alkarkhi AFM and Alqaraghuli WAA, Discriminant Analysis and Classification, *Easy Stat Food Sci with R* **1**:161–175 (2018).

41. Dray S and Dufour A, The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* **22**:1–20 (2007).

42. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2018). Available online: https://www.R-project.org/ (accessed on 02/01/2019).

43. Wu Z, Snyder G, Vadnais C, Arora R, Babcock M, Stensaas G, Doucette P, Newman, T, User needs for future Landsat missions, *Remote Sens Environ* **231**:111214 (2019).

44. Zeileis A and Hothorn T, partykit: A toolkit for recursive partytioning. *J Mach Learn Res* **16**:3905–3909 (2015).

45. Hothorn T, Hornik K, and Zeileis A, Unbiased recursive partitioning: A conditional inference framework, *J Comput Graph Stat* **15**:651–674 (2006).

46. Liaw A, Wiener M, Classification and regression by randomForest, *R News* **1**:18-22 (2002).

47. Genuer R, Poggi JM, and Tuleau-Malot C, Variable selection using random forests, *Pattern Recognit Lett* **31**:2225–2236 (2010).

48. Kuhn M, caret: Classification and Regression Training, *J Stat Softw* **28**:1-26 (2008).

49. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, and Ferreira LG, Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* **83**:195–213 (2002).

50. Jiang Z, Huete AR, Didan K, and Miura T, Development of a two-band enhanced vegetation index without a blue band, *Remote Sens Environ* **112**:3833–3845 (2008).

51. Rondeaux G, Steven M, and Baret F, Optimization of soil-adjusted vegetation indices, *Remote Sens Environ* **55**:95–107 (1996).

52. Huete AR, A soil-adjusted vegetation index (SAVI), *Remote Sens Environ* **25**:295–309 (1988).

53. Rouse JW, Haas RH, Schell JA, and Deering DW, Monitoring vegetation systems in the great plains with ERTS. In Third ERTS-1 Symposium, section A, NASA, Washington, DC, 1st ed., 309-317, (1973).

54. Vogelmann JE, Rock BN, and Moss DM, Red edge spectral measurements from sugar maple leaves, *Int J Remote Sens* **14**:1563–1575 (1993).

55. Deng W, Huang YB, Zhao CJ, and Wang X, Identification of seedling cabbages

and weeds using hyperspectral imaging, *Int J Agric Biol Eng* **8**:65–72 (2015).

56. Garcia-Ruiz FJ, Wulfsohn D, and Rasmussen J, Sugar beet (*Beta vulgaris* L.) and thistle (*Cirsium arvensis* L.) discrimination based on field spectral data, *Biosyst Eng* **139**:1–15 (2015).

57. Daughtry CST, Agroclimatology: Discriminating crop residues from soil by shortwave infrared reflectance, *Agron J* **93**:125–131 (2001).

58. Conforti M, Froio R, Matteucci G, and Buttafuoco G, Visible and near infrared spectroscopy for predicting texture in forest soil: An application in southern italy, *IForest* **8**:339–347 (2015).

59. White K, Walden J, Drake N, Eckardt F, and Settle, J. Mapping the iron oxide content of dune sands, Namib Sand Sea, Namibia, using Landsat thematic mapper data. *Remote Sens Environ* **62**:30-39 (1997).

60. GRDC - Grains Research and Development Corporation. Spray application manual for grain growers. Module 20. Target-selectable sprayers, how they work and set-up considerations. 2017. Available online: https://grdc.com.au/ data/assets/pdf_file/0024/234654/20grdcsm20targetselectablesprayers.pdf.pdf (accessed on 02/02/2019).

61. Hatfield JL, Gitelson AA, Schepers JS, and Walthall CL, Application of spectral remote sensing for agronomic decisions, *Agron J* **100**: S-117-S-131 (2008).

62. Cao Z, Yao X, Liu H, Liu B, Cheng T, Tian Y, *et al.*, Comparison of the abilities of vegetation indices and photosynthetic parameters to detect heat stress in wheat, Agr Forest

Meteorol **265**:121-136 (2019).

63. Wang W, Yao X, Yao X, Tian Y, Liu X, Ni J, *et al*., Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. *Field Crops Research* **129**:90–98 (2012).

64. Zhao B, Duan A, Ata-Ul-Karim ST, Liu Z, Chen Z, Gong Z, *et al.*, Exploring new spectral bands and vegetation indices for estimating nitrogen nutrition index of summer maize, *Eur J Agron* **93**:113–125 (2018).

65. Schwalbert RA, Amado TJC, Nieto L, Varela S, Corassa GM, Horbe TAN, *et al.*, Forecasting maize yield at field scale based on high-resolution satellite imagery, *Biosyst Eng* **171**:179–192 (2018).

66. Almeida J, Santos JA, Miranda WO, Alberton B, Morellato LPC, and Torres RS, Ecological informatics deriving vegetation indices for phenology analysis using genetic programming, *Ecol Inform **26***:61–69 (2015).

67. Suárez L, Zarco-Tejada PJ, Sepulcre-Cantó G, Pérez-Priego O, Miller JR, Jiménez-Muñoz JC, *et al.*, Assessing canopy PRI for water stress detection with diurnal airborne imagery, *Remote Sens Environ* **112**:560–575 (2008).

68. Wu C, Niu Z, Tang Q, and Huang W, Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation, *Agric For Meteorol* **148**:1230–1241 (2008).

69. Gabriel JL, Zarco-Tejada PJ, López-Herrera PJ, Pérez-Martín E, Alonso-Ayuso M, and Quemada M, Airborne and ground level sensors for monitoring nitrogen status in a maize crop, *Biosyst Eng* **160**:124–133 (2017).

70. Fern RR, Foxley EA, Bruno A, and Morrison ML, Suitability of NDVI and OSAVI

as estimators of green biomass and coverage in a semi-arid rangeland, *Ecol Indic* **94**:16–21 (2018).

71. Ren H and Zhou G, Estimating green biomass ratio with remote sensing in arid grasslands, *Ecol Indic* **98**:568–574 (2019).

72. Christensen S, SØgaard HT, Kudsk P, NØrremark M, Lund I, Nadimi ES, *et al.*, Site-specific weed control technologies, *Weed Res* **49**:233–241 (2009).

73. Hansen KD, Garcia-Ruiz F, Kazmi W, Bisgaard M, La Cour-Harbo A, Rasmussen J, *et al.*, An autonomous robotic system for mapping weeds in fields, *IFAC Proc Vol* **8**:217–224 (2013).

74. Stoner ER and Baumgardner MF, Characteristic Variations in Reflectance of Surface Soils1, *Soil Sci Soc Am J* **45**:1161 (2010).

75. Demattê JAM, Terra FS, Quartaroli CF, Spectral behavior of some modal soil profiles from São Paulo state, Brazil, *Bragantia* **71**:413–23 (2012).

Tables:

Table 1 The number of readings for each target evaluated in field trial and on-farm fields (5 total).

Target	Field trial	Field 1	Field 2	Field 3	Field 4	Field 5
Amaranthus hybridus	52	-	-	-	-	9
Avena sativa	-	-	-	-	9	-
Avena strigosa Schreb.	-	-	-	-	9	-
Bidens Pilosa	52	-	-	-	-	-
Brachiaria plantaginea	48	-	-	-	-	-
Brassica napus L.	-	-	-	-	9	-
Conyza bonariensis L.	-	9	9	-	-	-
Echium plantagineum L.	-	-	-	-	9	-
Euphorbia heterophylla	52	-	-	-	-	-
Glycine max	34	-	-	41	-	-
Ipomoea grandifolia	52	-	-	-	-	-
Lollium multiflorum Lam.	-	-	-	-	36	-
Panicum maximum	51	-	-	-	-	-
Polygonum convolvulus L.	-	9	-	-	-	9
Raphanus silvestris Lam.	-	-	-	-	9	-
Richardia brasiliensis Gomes	-	-	9	-	-	-
Sida rhombifolia	52	9	-	-	9	-
Solanum americanum Mill.	-	9	-	-	9	18
Sonchus oleraceus L.	-	-	-	-	9	-
Triticum aestivum L.	-	-	-	-	9	-
Total Weeds	393	36	18	41	117	36
Residues	25	-	-	50	-	-
Clay soil	-	20	10	-	-	-
Sandy soil	39	-	-	-	-	-

Field trial: Santa Maria; On-farm fields, Field 1: Não-Me-Toque; Field 2: Tapera; Field 3: Cruz Alta; Field 4: Júlio de Castilhos; Field 5: Itaara.

Table 2 Separation of wavelengths in spectral bands. Centered refers to the medium wavelength of each spectral band.

Band	Centered	Range
Coastal blue	425	400-450
Blue	480	450-510
Green	545	510-580
Yellow	605	585-625
Red	660	630-690
Red edge	725	705-745
Near infrared (NIR)	835	770-900

Table 3 Vegetation indices and their respective equations evaluated in this study.

Index	Equation	Reference
EVI	$2.5(R_{NIR} - R_{Red}) / (R_{NIR} + 6R_{Red} - 7.5R_{Blue} + 1)$	Huete *et al.* [49]
EVI2	$2.4 (R_{NIR} - R_{Red}) / (R_{NIR} + R_{Red} + 1)$	Jiang *et al.* [50]
OSAVI	$1.16(R_{NIR} - R_{Red}) / (R_{NIR} - R_{Red} + 0.16)$	Rondeaux *et al.* [51]
SAVI	$1.5(R_{NIR} - R_{Red}) / (R_{NIR} - R_{Red} + 0.5)$	Huete [52]
NDVI	$(R_{NIR} - R_{Red}) / (R_{NIR} + R_{Red})$	Rouse *et al.* [53]
NIR - RED	$R_{NIR} - R_{Red}$	Vogelmann *et al.* [54]
NIR / RED	R_{NIR} / R_{Red}	Vogelmann *et al.* [54]

R = reflectance.

Table 4 Statistics of training threshold when analyzed the validation data.

Index	Statistics	
EVI	Overall accuracy	98.5%
	Kappa	0.97
EVI2	Overall accuracy	98.1%
	Kappa	0.97
OSAVI	Overall accuracy	100%
	Kappa	1
SAVI	Overall accuracy	99.2%
	Kappa	0.98
NDVI	Overall accuracy	95.6%
	Kappa	0.93
NIR - RED	Overall accuracy	98.1%
	Kappa	0.97
NIR/RED	Overall accuracy	95.3%
	Kappa	0.93

Figures:

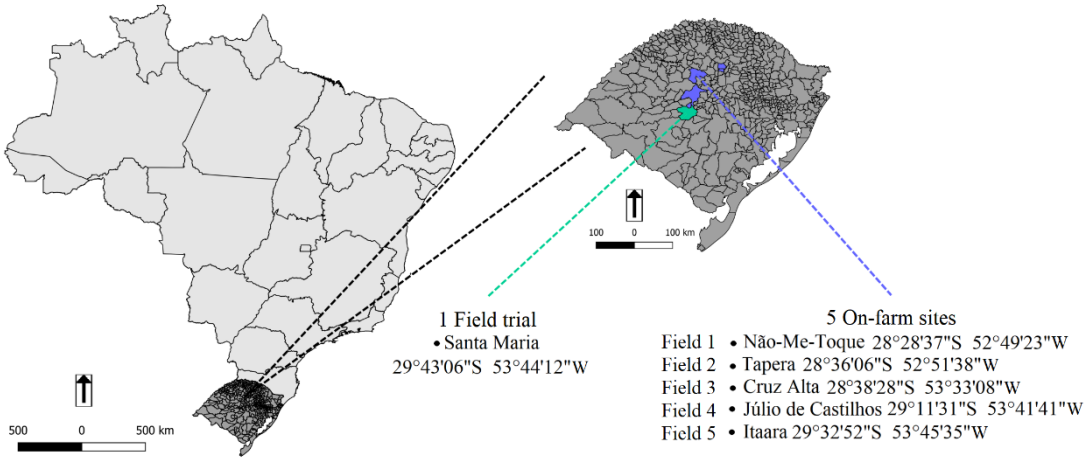

Fig. 1 - Field trials were located in the South of Brazil, the Rio Grande do Sul, comprising a training (one field study, Santa Maria) and validation data sets (five on-farm fields).

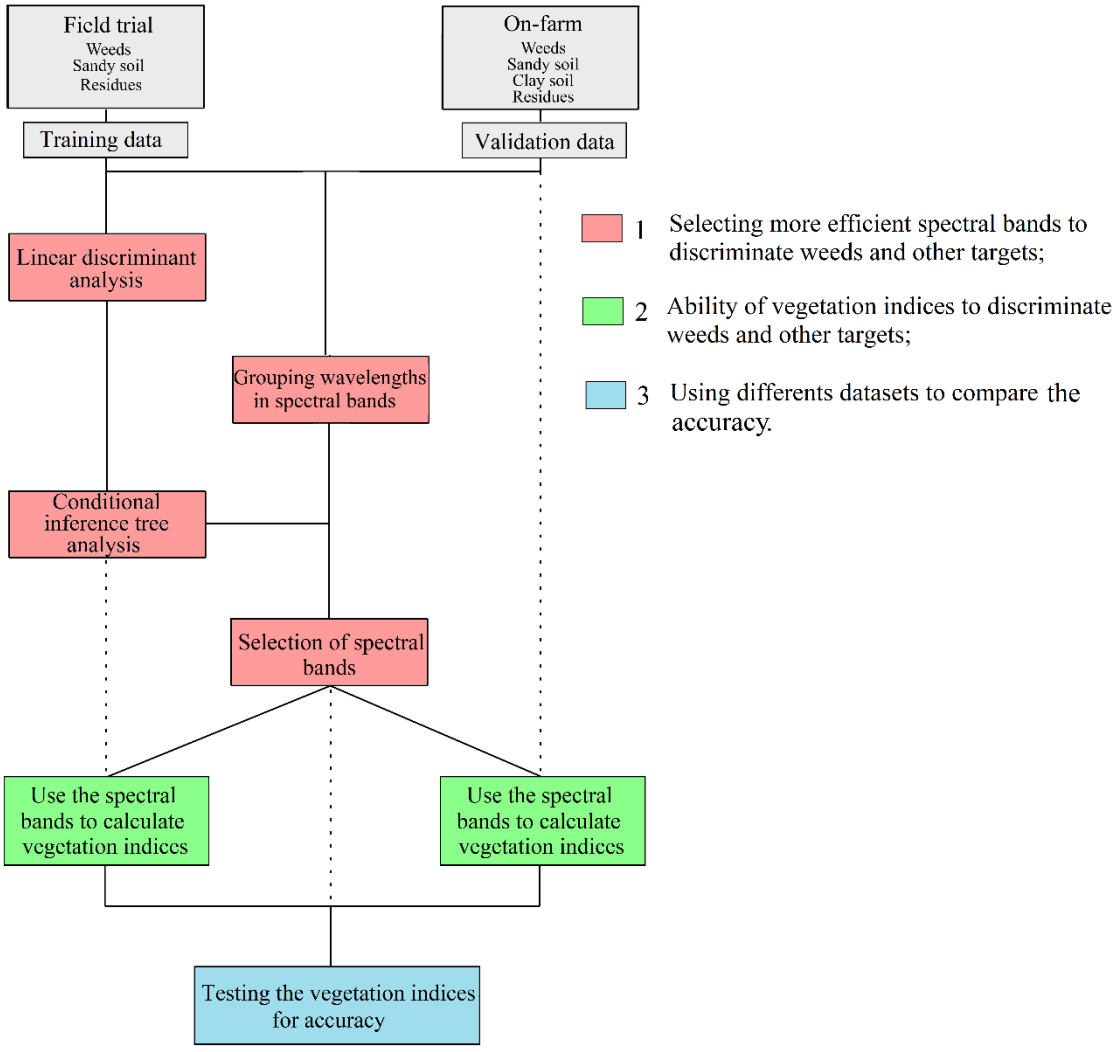

Fig. 2 - Theoretical framework for the study.

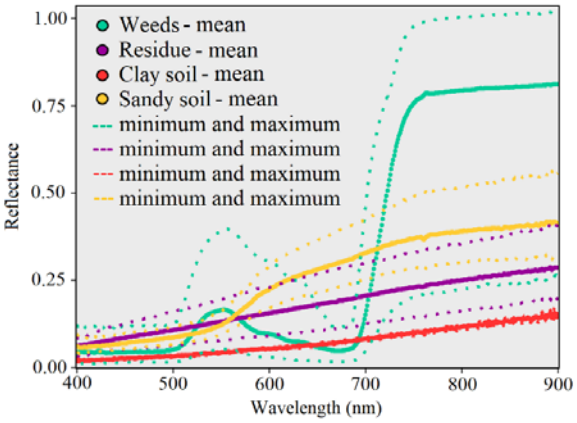

Fig. 3 - Spectral reflectance curves (mean, minimum and maximum) for weeds, residue, and soil type (clay, sandy) classes for the range of wavelength from 400 to 900 nm for the field trial.

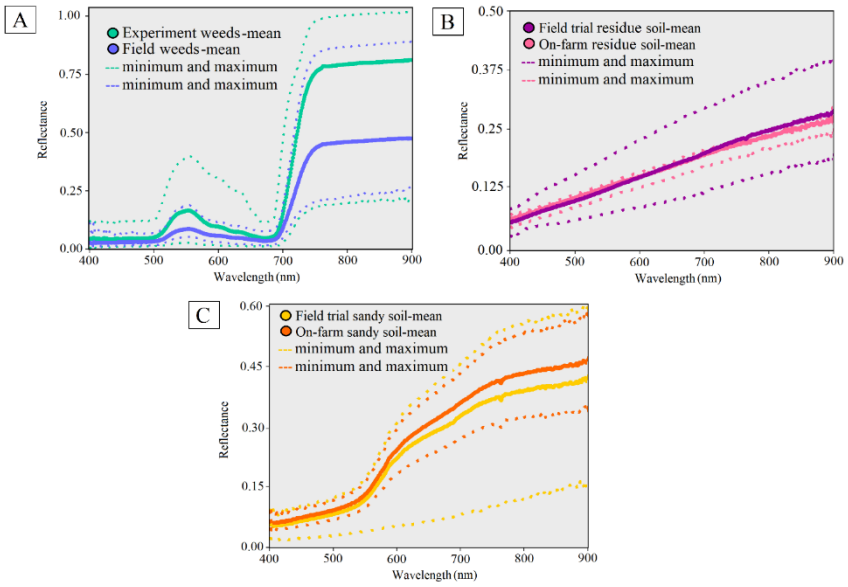

Fig. 4 - Comparison of training (data from one field trial) and validation (data from five on-farm sites) spectral curves, portraying minimum and maximum spectral curves, for the weeds class (A), residues (B), and sandy soil (C).

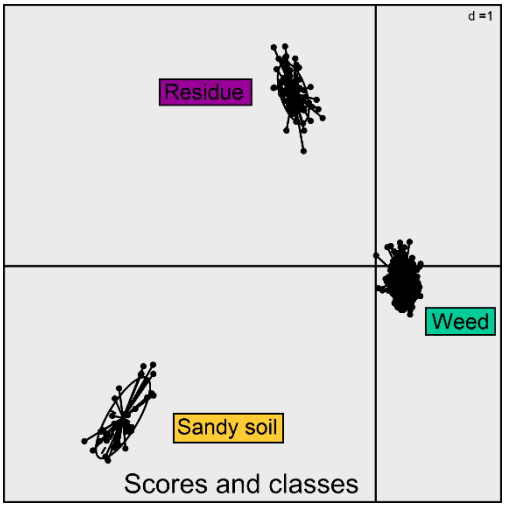

Fig. 5 - Linear discriminant analysis (LDA) for all the classes evaluated in this study for the field research (residue, sandy soil, and weeds). Ellipses represent the confidence regions around the mean of canonical scores at the 95% confidence level.

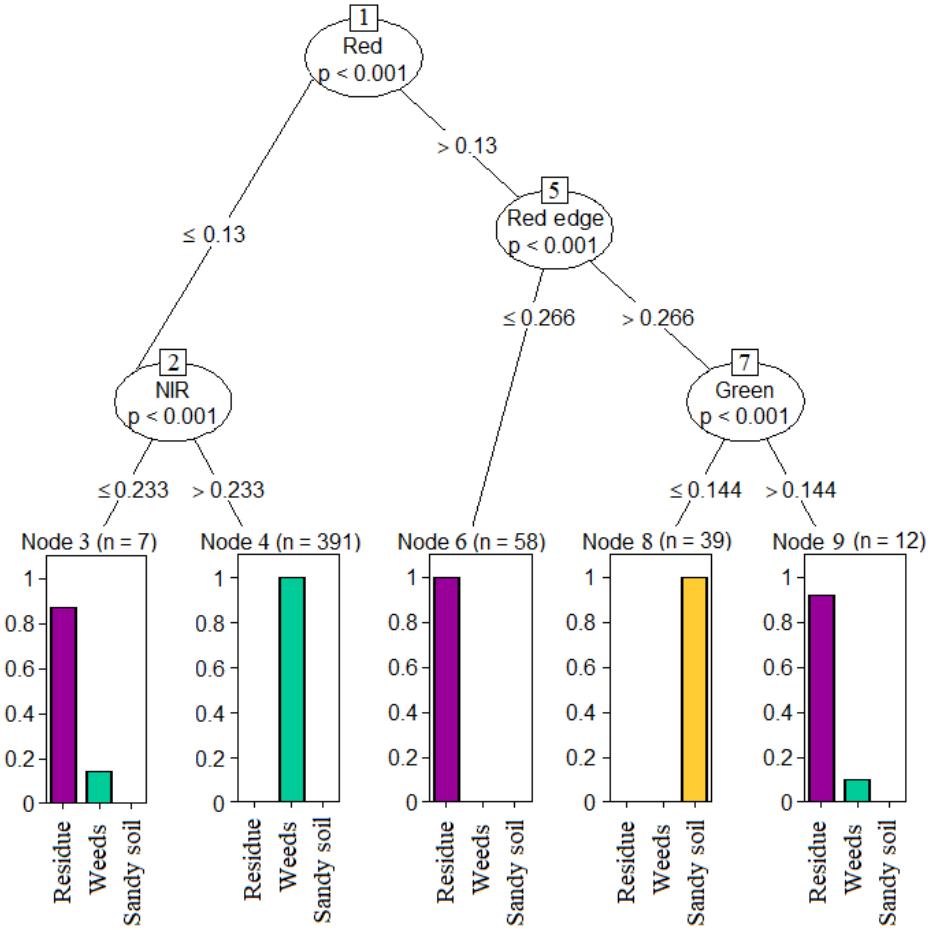

Fig. 6 - Conditional inference tree analysis using the bands: coastal blue, blue, green, yellow, red, red edge, and near-infrared. Bars at the bottom of the figure represent the density of points for each target. The values of conditions for classification are reflectance values.

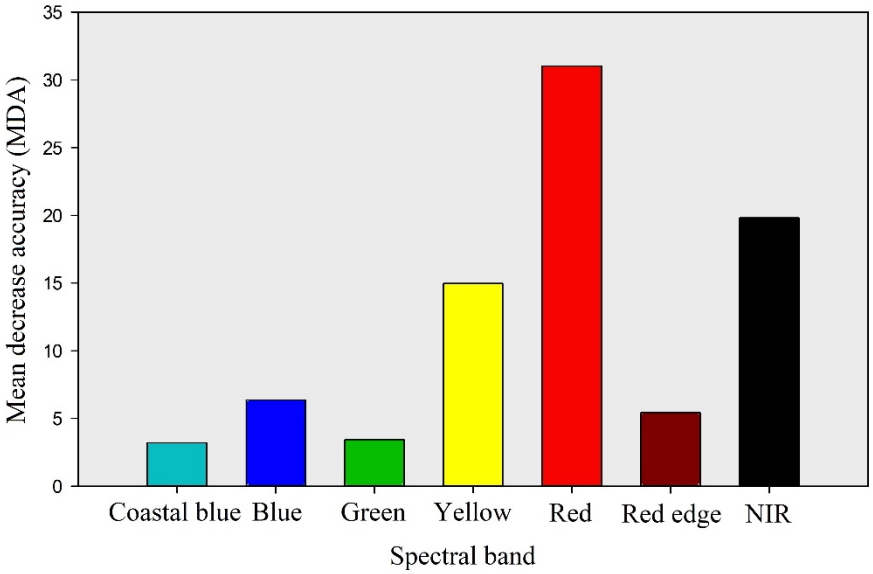

Fig. 7 - Mean decrease accuracy (MDA) of the spectral bands utilizing training and validation data in random forest analysis.

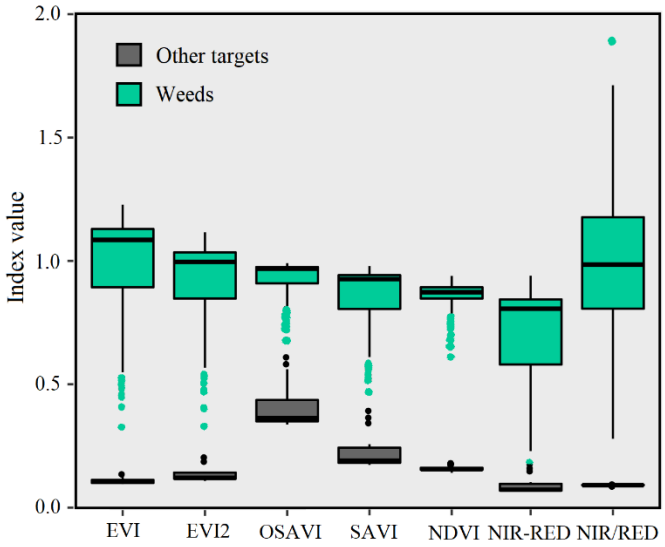

Fig. 8 - Boxplot representing data distribution for different bands and indices for training data. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are considered outliers. NIR/RED values were divided by 15 to better fit in y axis.

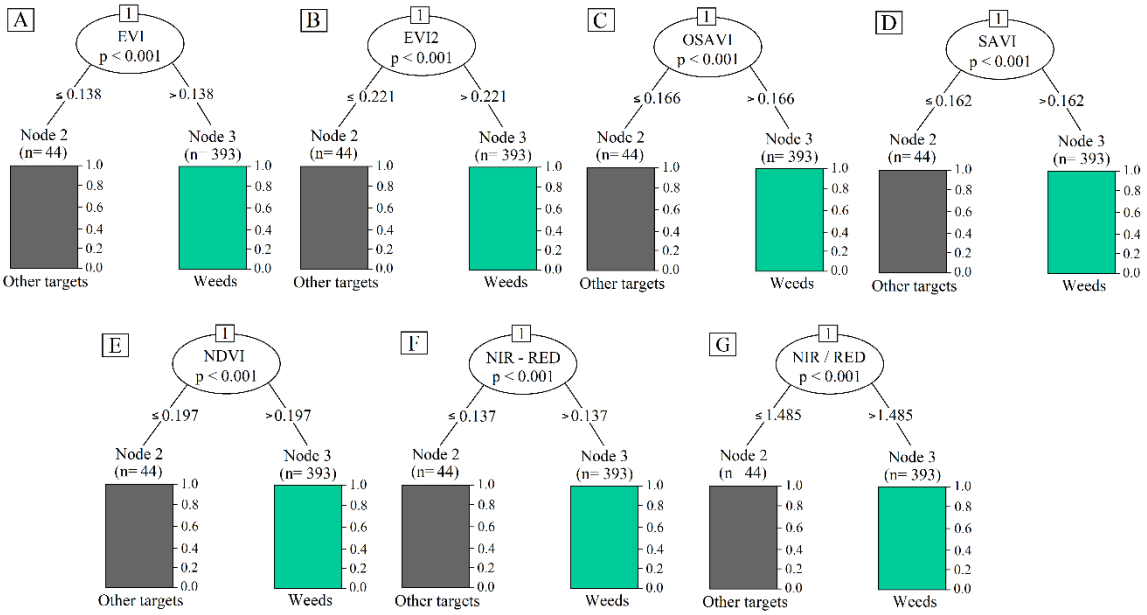

Fig. 9 - Condition inference tree of EVI (A), EVI2 (B), OSAVI (C), SAVI (D), NDVI (E), NIR – RED (F) and NIR/RED (G).