Use of supervised and unsupervised approaches to make zonal application maps for variable-rate application of crop growth regulators in commercial cotton fields

Maria Carolina da Silva Andrea
maria.carolina@nuvem.net

Nuvem Tecnologia https://orcid.org/0000-0001-9633-6622

Cristiano Ferreira de Oliveira
Nuvem Tecnologia

Fabrícia Conceição Menez Mota
Nuvem Tecnologia

Rafael Carlos dos Santos
Nuvem Tecnologia

Edilson Freire Rodrigues Junior
Nuvem Tecnologia

Lucas Monteiro Bianchi
Nuvem Tecnologia

Rodrigo Souza Oliveira
Nuvem Tecnologia

Caio Marchiorato de Gouveia
Nuvem Tecnologia

Victor Gabriel Souza Barbosa
Nuvem Tecnologia

Marco Antônio Bispo e Silva
Nuvem Tecnologia

Research Article

Keywords: cotton, site-specific management, growth regulator, Cerrado

Posted Date: July 15th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4522048/v1

Version of Record: A version of this preprint was published at Journal of Cotton Research on January 6th, 2025. See the published version at https://doi.org/10.1186/s42397-024-00204-y.

Use of supervised and unsupervised approaches to make zonal application maps for variable-rate application of crop growth regulators in commercial cotton fields

Maria Carolina da Silva Andrea[1]*; Cristiano Ferreira de Oliveira[1]; Fabrícia Conceição Menez Mota[1]; Rafael Carlos dos Santos[1,2]; Edilson Freire Rodrigues Junior[1]; Lucas Monteiro Bianchi[1]; Rodrigo de Souza Oliveira[1,3]; Caio Marchiorato de Gouveia[1,4]; Victor Gabriel Souza Barbosa[1,5]; Marco Antônio Bispo e Silva[1]

*Correspondence: maria.carolina@nuvem.net; [1]Nuvem Tecnologia, Cuiabá, MT, Brazil; [2]Institute of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil; [3]Institute of Math and Computer Science, University of São Paulo, São Carlos, SP, Brazil; [4]Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, MG, Brazil; [5]Department of Computer Science, Federal University of Paraná, Curitiba, PR, Brazil.

Abstract

Background

Zonal application maps aim to represent field variability by means of variables of interest that can further be translated into management practices. Zonal application maps for crop growth regulators in cotton under variable-rate strategies are commonly based exclusively on vegetation index variability. However, saturation by vegetation indexes cannot be avoided by multispectral imagery and dense crop vegetation areas. The objective of this study was to compare zonal application maps for crop growth regulator application under variable-rate conditions via two approaches: (i) relying on field-measured crop height and (ii) not relying on field-measured crop height. During the agricultural season, we developed zonal application maps using an unsupervised framework, representing local variability based on field-collected data, satellite imagery data, soil texture and phenology. Posteriorly, using data from 3 agricultural seasons, we developed a supervised framework to predict plant height by using a machine learning algorithm based on remote sensing and phenology data to test the development of the same maps without field data on plant height.

Results

The results indicated good performance of the machine learning model for predicting plant height, but these predictions presented much lower variability than that found in field conditions. Ultimately, the predictions went through the same unsupervised process to construct maps, but without requiring field-measured data. We tested both approaches in three fields on two different dates each. Fields with intense textural variability provided the highest compatibility between the maps of both approaches. Conversely, the lowest compatibility was found in fields with lower soil textural variation.

Conclusion

There is potential for the use of predictive modelling to assist in the construction of zonal application maps, but the adherence to real patterns of crop growth variability found in the field is highly variable and must be assessed on a case-by-case basis.

Keywords: cotton; site-specific management; growth regulator; Cerrado

1. **Introduction**

Brazil is one of the world's main cotton producers. In terms of volume, it is behind China and India and is very close to U.S. production (OECD, 2023). In the 2022-2023 agricultural season, Brazil had a cultivated area of 1663 thousand hectares, with an expected increase of 5% in area for the 2023-2024 harvest. During 2022-2023, the gross productivity was 4257 kg/ha (average lint yield of 41%). At the national level, the Midwest region stands out: during 2022-2023, the macroregion had a cultivated area of 1291 thousand hectares and an average gross productivity of 4588 kg/ha. The largest national producing state, Mato Grosso, had a cultivated area of 1190 thousand hectares (95% of the Midwest region and 71% of the national territory) and an average gross productivity of 4580 kg/ha (CONAB, 2023). In Mato Grosso, and most of the Midwest region, where much of the country's agricultural activity is located, the predominant biome is the Cerrado (Brazilian savanna). Areas dedicated to agricultural activity in the Cerrado are characterized mainly by weathered soils, high acidity and low natural fertility. In general, these areas are very wide and flat, facilitating the presence of large-scale mechanization and production units. Cotton production in the Cerrado is a high-value but also high-risk activity considering the high cost of inputs and the risk associated with environmental conditions. Agricultural production in these areas is essentially rainfed, and the rainy season in most of the state extends between October and April, with average annual totals between 1200 and 2000 mm (Souza et al. 2013), where the highest accumulated values are closer to the southern border of Amazonia. Cotton is a crop with unique aspects of cultivation. It is a perennial plant cultivated annually, and to consistently achieve the highest possible yields in each season and enable mechanization, its vegetative and reproductive growth must be well balanced through efficient crop management (Silvertooth et al. 1996). Environmental factors (e.g., temperature, cloudiness, rainfall) and management (such as nitrogen application) easily interfere with this balance. Therefore, crop growth must be frequently monitored, and interventions with the Crop Growth Regulator (CGR) are often necessary. The use of CGRs became common practice in the management of this crop. Field assessments of plant height and certain plant structures, such as the number of nodes, continue to constitute some of the most commonly information used for crop monitoring (Echer and Rosolem, 2022). In the Brazilian Cerrado, it is usually recommended that the first application of CGR occur approximately 35 days after emergence (dae), and the last occurs approximately 115 dae, which is responsible for the total stoppage of vegetative growth. The volume of product, the number of applications and respective quantities are locally defined to meet such planning and the local needs of the plant (Berger et al., 2019).

As with many inputs in the production cycle, it is recommended that the CGRs be applied according to site-specific needs within the field (instead of an average value) in the context of precision agriculture (PA) (Plant, 2001). This is especially important when we consider large commercial production areas (fields with 100+ ha) that invariably exhibit spatial variability in environmental conditions, such as soil texture, that ultimately reflects in plant development. Site-specific management measures for crop inputs, which can be enabled by variable rate (VR) techniques of application, are essential for achieving more sustainable production, more efficient use of products, and maximization of crop profitability levels, the core of PA (Plant, 2001; Taylor and Fulton, 2010). VR applications in PA usually follow one of two paths: map-based or sensor-based systems. In the map-based system, the entire process is relatively slower: data are collected and later analysed through an algorithm, and the product (zonal application map) is returned to the user (Taylor and Fulton, 2010). Although sensor-based systems are faster (involves real-time measurements and processing), currently, their use is limited by their high cost for small producers and for large producers when considering the number of machinery necessary for this type of large-scale application. This makes the map-based systems an important part of the process of adopting PA techniques in large-scale commercial fields.

Many sensor and map-based systems are based on the plant's reflectance response, such as through the use of vegetative indices (VIs). VIs, including the well-known Normalized Difference Vegetation Index (NDVI), are simple algorithms that offer information about the spectral signatures of vegetation in terms of structure, physiology and biochemistry (Janse et al., 2018). These methods are commonly employed in precision agriculture and crop modelling studies to determine the connection between vegetation and radiation obtained from remote sensing. The relationship between the VI and the plant trait is an indirect one, as the VIs are proxies between these traits. It is important to highlight that the NDVI (as most VIs at some point) tends to saturate under medium to high biomass conditions, despite its widespread utilization throughout plant development. Under conditions of medium to high vegetation density, which is often the case for row crops, this phenomenon is already evident at a certain stage of plant development (Gitelson, 2004; Payero et al. 2004; Gutierrez et al. 2012). Plant height is the most common biophysical characteristic evaluated in field monitoring and intimately related to the understanding of cotton's necessity for crop growth regulator application (Taylor and Fulton, 2010). The relationships of remote sensing bands and VIs with plant height are also widely addressed and known (Payero et al. 2004). Although some of these variables may present a linear association that has been widely used to direct the use of this input, this relationship is present only at the beginning of the crop cycle. At a certain phenological stage the VI saturates, providing no further information about the continuity of plant growth, and representing a limitation in its exclusive use for growth regulator management (Taylor and Fulton, 2010). The data used in the present study were primarily evaluated with respect to the NDVI saturation associated with plant height and plant phenology (Figure 1). With field data, it was possible to confirm the saturation in the height versus NDVI relationship and the phenological stage at which this occurred. Such information is important for planning management, as it clarifies at which point in the crop cycle the use of IVs becomes less useful for certain applications and was a starting point for the present study.

[Figure 1]

Figure 1. Relationships between days after emergence and the NDVI of cotton fields (left figure) and between the measured cotton height and the NDVI across agricultural seasons (right figure) of the fields assessed in the present study.

It is common in PA to use management zones (MZ) to guide the site-specific application of crop inputs. The MZ, by definition, is an area in which the expression of crop development is relatively homogeneous and is associated with limiting factors, which can be managed uniformly. The factors by which MZ are usually delimited can be quantitative and stable (topography; physical and chemical-biological characteristics of the soil, such as texture, organic matter, and pH); quantitative and dynamic (productivity, soil moisture and salinity, and soil N status); qualitative and stable (soil survey maps; immobile nutrients such as P); and intuitive and historical (farmers' knowledge, historical practices, and crop rotation). The inputs for which these MZ are used to guide applications are usually N, P, K, lime, seeds and cultivars (Doerge, 1999). In general, MZ are used to represent the productive potential of areas and can change considerably from year to year (Plant, 2001) or remain relatively stable when delineated based on a longer history and understanding of the area (Santi et al., 2016; Damian et al. 2020). Within-field areas allocated to site-specific management of an input, a common implementation of PA, and their inter- and intra-annual variability can be associated with climate variability (Alesso and Martin, 2023), as also with the management of dynamic factors in the field, such as pests and diseases (Méndez-Vazquez et al. 2019). There are difficulties in using the traditional PA approach to delineate MZ through the use of only temporally stable factors to manage dynamic occurrences, and a possible solution to this is the inclusion of dynamic information layers (Méndez-Vazquez et al. 2019). Therefore, we identify the oppportunity for studies that adapt the concept of MZ to represent dynamic characteristics of the field throughout the season (Pierce and Nowak, 1999), such as the biometric properties of the plant, and that go beyond the use of VIs and their inevitable saturation.

A final relevant point in developing zonal application maps is the algorithm. They can be ready-to-use algorithms (e.g., obtained by research institutions), but it is important to consider local adaptation of methodologies, as this can captures local conditions and specificities (Liu et al., 2021; Vaz et al., 2023). Most studies use unsupervised multivariate techniques such as clustering techniques and principal component analysis (PCA) to delineate management zones (Fridgen et al., 2004; Córdoba et al., 2016; Damian et al., 2020; Ouazaa et al., 2022; Vaz et al., 2023). One of the main objectives of principal component analysis (PCA) is to reduce the dimensionality and retain the maximum possible variance of the variables, a task performed by creating a linear combination of conceptually relevant variables. The application of PCA in agriculture has been highlighted both by authors in the statistical field (Jolliffe, 2002) and by authors in the application areas, as in precision agriculture techniques (Córdoba et al., 2016; Ortuani et al., 2019; Ouazaa et al., 2022; Vaz et al. 2023) for different crops and aspects of their management.

The present study represents a continuation of a study (Andrea et al. 2023). The most relevant points of this study are as follows:

- Large commercial fields exhibit high variability in terms of soil texture and consequently crop development;
- A large number of prescription maps for CGR use are based solely on VIs; however, local saturation of the VI (when relating to crop height) occurs approximately at 75 days after emergence;
- An unsupervised framework (UF) was developed to make zonal application maps under demand during the agricultural season based on field variability in crop height measurements, soil texture and vegetation indices;
- A supervised framework (SUF) was used as an alternative to field measurements of crop height, after which the previous unsupervised analysis was performed to create the zonal application maps,
- The zonal application maps for VR applications from both frameworks were compared in terms of their compatibility and their relationship with the input variables;

The objective of this study was to compare and assess zonal application maps for CGR management under VR application via two approaches: (i) relying on field-measured crop height and using an unsupervised approach to define the application maps (UF) and (ii) not employing field measurements and using a combination of supervised and unsupervised approach (SUF) to develop the zonal application maps.

2. Materials and Methods

2.1. Study area

In this study, data from approximately 200 fields (Table 1) located on three commercial farms located in Mato Grosso state, Brazil, and owned by the same company, were used. Approximately 100 thousand hectares of commercial planted area are planted on the farms, and the main business is based on cotton and soybean production. Crop management practices, including soil and plant sampling, fertilization and plant health (preventive and treatment), are carried out on all farms following the same protocol. The main differences are therefore due to the environmental characteristics of each location (meteorological and soil conditions). Despite the data from the various plots mentioned above, they refer to the usual field observations that are part of the farm's management routine. The number and arrangement of sampling points are heterogeneous (in space and time) and may be considered less than ideal when compared to experimental conditions. This is due to the size of farms and plots and the labor intensity of such activities. The application maps were constructed for three of these fields (Table 2), which were previously selected and in accordance with the local field agronomists, where there was the possibility of more frequent and homogeneously distributed field sampling during the agricultural season.

Table 1. General information about all fields that contributed to the supervised modelling dataset.

Farm	Fields	Texture range of all fields
1	65	sandy (15%) - medium-clayey (25-40%)
2	35	sandy (15%) - clayey (40-60%)
3	108	sandy (15%) - clayey (40-60%)

Obs: Table 1 refers to a large quantity of fields for which often only a few plant height measurements were obtained. Therefore, these areas will not be detailed in this study and were used only to provide observations for the predictive model. The number of fields represents the average of fields evaluated in each agricultural season. Field samples were evaluated throughout three seasons (2020-2021, 2021-2022, 2022-2023).

Each recorded observation (in a given geographic position, latitude and longitude) was associated (by nearest distance) with its corresponding point of information of soil texture and remote sensing, along with days after emergence, representing an observation for predictive modelling. Crop height measurements were performed across each growth cycle, totalizing approximately 30 thousand georeferenced points in all three agricultural seasons.

2.2. Input data

Field data: crop height, soil texture, days after emergence

The data used to construct the modelling dataset (Table 1) are representative of several fields, with variable monitoring frequency during each harvest, as they represent the farm's usual commercial activities. In some of the plots for which zonal application maps were constructed, there were more frequent and homogeneously distributed measurements during the agricultural season. For these fields, a map comparison assessment concerning the 2022-2023 agricultural season was carried out. The fields were between 100 and 250 hectares in size and are distributed across flat areas (up to 3% slope and medium-clayey soil texture) and marginal areas (more intense soil texture and slope variability). The characteristics of the fields assessed in greater detail can be found in Table 2.

Table 2. Data on cultivation in the fields that were closely monitored and for which application maps were developed.

Farm	Field	Soil Texture[1]	Emergence date[2] (mm/dd/YY)	Plant stand count[3]	CGR applic.[4]	Accum. rainfall (mm)	Harvest date[2] (mm/dd/YY)	Final yield (@ ha^{-1})[5]
1	A	medium-clayey (25-40%)	01/02/23	10.5	7	801.8	07/13/23	256,25
2	B	medium-clayey (25-40%)	01/17/23	9	13	1001.0	07/26/23	307.04
3	C	medium-clayey (25-40%)	02/10/23	10.9	7	759.2	08/29/23	242,04

[1] Refer to the average soil texture of the entire field; [2] Emergence and harvest date of the respective fields during the 2022-2023 agricultural season; [3] Number of plants per meter in the rows; [4] Total number of crop growth regulator applications (variable and uniform rate); [5] Average yield of each field, where 1 @ = 15 kg. Cotton cultivars: A: FM 911 GLTP C2; B: FM 944GL C1; C: TMG22GLTP S2.

The fields that were closely followed had an average of 30 points of measurement per event of monitoring, which occurred at a frequency of approximately every ten days, although these values could vary between and within farms due to events such as heavy rain, application of pesticides and dynamics of planning other operations. The registered information (plant height and their respective coordinates) represented the average of at least three nearby plants, and the height was measured using a manual instrument (calibrated ruler). In Figure 2, an example of one field and the sampling distribution in one of the measurement events are shown.

[Figure 2]

Figure 2. Distribution of plant height measurement points in Field A during the 2022-2023 agricultural season.

Soil texture data were obtained from the farm's database and refers to its usual soil sampling practices. Soil dataset was a grid with a spatial resolution of 10 m. Days after emergence (dae) was the variable used to represent plant phenology. These data also come from farm records regarding the day of emergency in each field.

Remote sensing data and dataset

Satellite multispectral imagery from Planet, the Planetscope satellite constellation (consisting of 180+ CubeSats), was used in this study. The images obtained for all fields consisted in a 4-multispectral band product covering the visible and near-infrared range, with the following spectral ranges: blue (455 - 515 nm), green (500 - 590 nm), red (590 - 670 nm) and NIR (780 - 860 nm). The Planetscope products represented surface reflectance with 16-bit depth and 3-m pixel size and are corrected for geometric and atmospheric effects (6SV2.1 radiative transfer code). More details about the products can be found in the Planet imagery product specifications (Planet, 2022). This data has daily temporal resolution, although in practice this may change specially between October and March due to the rainy season, since not all images can be used due to rain and heavy cloud cover. The images used in this study were available to the end-client at the specific shape of each commercial field and were specifically acquired for fields with cotton cultivation; therefore, disturbances related to other types of vegetation, such as weeds, were minimal to none. The remote sensing data (bands and VIs) used in the present study can be found in Table 3.

Table 3. Summary of the remote sensing bands and vegetation indices used in the present study.

Index/band	Description	Equation	Reference
R, G, B, NIR	Red, Green, Blue, Near-Infrared bands	-	-
NDVI	Normalized Difference Vegetation Index	$\frac{(NIR - R)}{NIR + R}$	Verrelst et al. 2019
DVI	Difference Vegetation Index	$NIR - R$	Verrelst et al. 2019
GNDVI	Green NDVI	$\frac{(NIR - G)}{NIR + G}$	Gitelson et al. 1996
MSAVI2	Modified Secondary Soil-Adjusted Vegetation Index	$\frac{2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR}}{2}$	Mróz and Sobieraj, 2004

The choice for these specific bands and vegetation indices as shown in Table 3, follows previous study (Andrea et al. 2023), while a variability of these products allows for a broader assessment of crop traits, phenological phases (Gutierrez et al. 2012) and environmental conditions (soil, atmospheric effects and others) (Payero et al. 2004; Xue and Su, 2017). Most vegetation indices use the NIR and R bands (except for the GNDVI), specifically due to their well-stated sensitivity to photosynthetically active biomass (Tucker, 1979; Gutman et al. 2021). On days when crop height was measured in a specific field, a corresponding satellite image associated with that field was used (with a maximum interval of up to 5 days from field sampling) to compose the dataset of the respective field.

2.3. General framework

In Figure 3 is possible to observe the main steps carried out in both approaches of this study. While the input data of the UF and SUF approaches are almost the same, their use is ultimately linked with the availability of field data on plant height and the use of predictive modelling. Further details of the steps carried out in UF and SUF can be found in the following subsections.

[Figure 3]
Figure 3. Flowchart summarizing the methodological approaches used in this study.

UF-PCA-based approach

The approach taken in UF started with some of the fields assessed for the 2022-2023 agricultural season (1). The data from each of these fields (soil texture, vegetation index and height samples collected in the field) went through a manipulation process (2), which included: the interpolation of crop height measured in the field and the use of a spatial join algorithm to construct the dataset (3). A radial basis function (RBF) algorithm with a thin plate spline kernel was used to interpolate the plant height sample data. The algorithm is known for its quality of the results for producing meshes (Sastry et al. 2015), especially under conditions of scatter data in the environmental sciences (Keller and Borkowski, 2019). Plant height sample data had a relatively low spatial resolution (usually 30 points in fields with variable sizes) compared to remote sensing data (3 m). Because the soil texture data was already at 10 m of spatial resolution, all the input data were placed at this same resolution to form each field-date dataset (step 3). To join all data, a core set KDTree (or cKDTree) algorithm was used, which builds an index for all the data and uses this structure for nearest neighbour queries. More details about this algorithm can be found in Narasimhulu et al. 2021. The interpolation and spatial join algorithms were applied through the *use of the SciPy* library (https://zenodo.org/records/6940349), and all the analysis were carried out using the python programming language.

To each of these datasets, an unsupervised, named principal components (PCs) approach was taken (4, 5) ultimately resulting in the zonal application maps (6). This approach represented the core of the UF algorithm. This analysis aimed to summarize (describe the data in fewer concepts than the original data) and reduce (derive factor scores from the factors that replace the original values) the information found in the field. The choice of technique was also based on the assumption that there is an underlying structure in the set of chosen variables (Jolliffe, 2002), resulting from the analysis of empirically observed relationships and patterns. In general, the process began by centralizing the data in a coordinate system, followed by the determination of the first best-fit line drawn on the mass of observation points (its explanatory power is directly related to the correlation of the data). This line, which represents the first PC or first latent variable, also goes in the direction of the maximum variance of the projections of the observations that fall on it. Ninety-degree projections are made of each observation on the line, and the distances from the origin to these projections are called scores (therefore, each observation has a score associated with each PC). Each PC has (i) a direction K-dimensional vector, also called loadings, which defines the best-fit line and represents a link with the latent variable of the coordinate system, and (ii) a score vector, which is a set of N (N = number of observations) score values along the best-fit line. The same process was performed for the second PC. The direction of this new PC also gives the greatest variance in the score values when projected onto this new vector. The use of two PCs, as in the present study, defined a plane that represented the latent variable model and was considered here the best representation of the data (Jolliffe and Cadima, 2016). To achieve these results, the following steps were taken, as demonstrated by equations 1 to 9: (i) data normalization; (ii) calculation of the correlation matrix; (iii) Bartlett sphericity test; (iv) determination of eigenvalues and eigenvectors; (v) determination of shared variance; (vi) determination of the communality of each variable; (vii) determination of the factorial score matrix; (viii) determination of factorial score variables based on factorial scores; (ix) determination of the factorial weighted variable, a variable that summarizes the observations weighted by the factorial scores; and (x) classification of the newly created variable.

(i) *Data normalization:* Data were normalized following the min–max procedure, as in Equation 1:

$$(1)\quad x' = \frac{x - min(X)}{max(X) - min(X)}$$

where x' = the normalized data point; x = the original data point; and $max(X)$ and $min(X)$ = the maximum and minimum observed values, respectively.

(ii) *Correlation matrix:* The correlation matrix was used for further eigenvalue extraction. The correlation matrix was determined by the Pearson correlation, as in Equation 2:

$$(2)\quad r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \bar{x})^2)(\sum_{i=1}^{n}(y_i - \bar{y})^2)}}$$

where r_{jk} = Pearson correlation coefficient (between -1 and 1); x_i and y_i = individual data points of two variables being compared; and $\underline{\bar{x}}$ and $\underline{\bar{y}}$ = mean values of the two variables.

(iii) *Bartlett's sphericity test:* The test performed on the correlation matrix assesses the viability of using the PCA (H0: correlation matrix is similar to an identity matrix; therefore, it is uncorrelated and not suitable for PCA) and can be determined through Equation 3:

$$(3)\quad \chi^2 = -\left[\frac{(n-1)-(2p+5)}{6}\right] log_e|R|$$

where χ^2= chi-square test statistic; n = sample size; p = number of variables with degrees of freedom df = 0.5 (p) (p-1); and |R| = determinant of the correlation matrix.

(iv) *Eigenvalues and eigenvectors:* The eigenvalues and eigenvectors are extracted based on the correlation matrix via the following general formula:

(4) $|A - \boldsymbol{\lambda} I|v = 0$

The equation is solved as follows: set the determinant $|A - \boldsymbol{\lambda} I|$ to zero to find the eigenvalues $\boldsymbol{\lambda}$; for each eigenvalue, solve equation $|A - \boldsymbol{\lambda} I|v = 0$ to find the corresponding eigenvector v.

(v) *Shared variance:* Determination of the shared and cumulative variance, as in Equation 5.

(5) $S = \lambda_k / \Sigma\, \lambda_i$

where S = shared variance; λ_k = eigenvalue of the kth PC; and $\Sigma\lambda_i$ = sum of all eigenvalues (total variance in dataset).

(vi) *Communality:* Indicates each variable's part of the variance shared with the other variables that is due to the common factors, as in Equation 6:

(6) $h_j = \sum_{j=1}^{p} s_{ij}^2$

where h_j = communality of the jth variable and s_{ij} = loading (or correlation) between the ith component and the jth variable.

(vii) *Factorial scores:* Determination of the factorial score matrix, as in Equation 7:

(7) $f_j = \left(\frac{v_{ij}}{\sqrt{\lambda_j^2}} \right)$

where f = the factorial score of the j^{th} factor, $\boldsymbol{\lambda}$ = the eigenvalue, and v = the eigenvector.

(viii) *Factorial score (vector) variables:* Determination of factorial score variables, as in Equation 8:

(8) $F_p = \sum(f_p \times p_i)$

where F_p = the component score vector of each p variable; f_p = the factorial score of the p variable; and p_i = the ith observation of the p variable.

(ix) *Component weighted variable:* Determination of the variable that summarizes the weighting of all variable observations by the chosen components, as in Equation 9:

(9) $F_w = \sum(F_p \times Sf)$

where F_w = the created component weighted variable, which summarizes the information obtained by the component load variables for the desired number of components by summing the association between the factorial score variable and the factors' shared variance.

(x) *Classification*: In this step, the F_w variable, representing the majority of the variance in the data weighted by the chosen PC scores, was classified using percentiles to construct a map containing 3 classes in the zonal application maps.

SUF: Predictive modelling approach

2.4.2. Without collected crop height: supervised and unsupervised frameworks

The framework concerning the supervised or SUF approach, as in Figure 3, initially presented the construction of a dataset of a large number of fields based on the previous three agricultural seasons (1). The dataset was manipulated (2) to contain all plant height information measured in the field, registered under a certain latitude and longitude, associated with days after emergence and with remote sensing information (by proximity to crop height measurements). This process formed the dataset for modelling (3), from which the predictive modelling approach took place (4, 5). In the modelling dataset, height was the dependent variable, and remote sensing products and days after emergence the independent variables. 85% of the data were split into train and validation sets (70% and 15%), and the remainder were used for test. In terms of model adjustment, a hyper parameter tuning process was carried out to improve the model's performance using a Bayesian optimization technique. For such process, some of the most common options include grid-search (or exhaustive search), random-search and Bayesian optimization search algorithms. The latter has mechanisms to overcome the widely reported problems of efficiency (compared to the use of grid search) and precision (compared to the use of random search). In this study the Bayesian Optimization process was used through a Python library (Nogueira, 2014). In general, this process uses a probabilistic model (commonly Gaussian processes) that approximates the true performance of the objective function, such as the mean squared error, as used in the present study. The search for the configuration of hyper parameters is carried out using an acquisition function (or exploration strategy), such as the upper confidence bound or expected improvement, balancing exploration and exploitation and placing more emphasis on more promising regions. More details about how the Bayesian optimization algorithm works can be found in Wu et al. 2019 and Snoek et al. 2012.

The predictions were evaluated using the following regression metrics: R-squared (R-squared), mean absolute error (MAE) and maximum absolute error (MAEr) (Botchkarev, 2019) (equations 10-12).

(10) $$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

(11) $$MAE = \frac{1}{n} \times \sum_{i=1}^{n} |y_i - x_i|$$

(12) $$MAEr = max_{i=1,n} |(y_i - x_i)|$$

where x_i = predicted at the i^th value; y_i = actual at the i^th value; and $\underline{y}$ = mean of the true values.

After the development of the supervised algorithm, it was used in datasets from the fields evaluated in the present study to predict plant height at dates later than the period covered in the modelling dataset. To this resulting dataset, soil texture data was added, and the same unsupervised process as previously described in UF (PC approach) was carried out. At the end of this process, the zonal application maps (6) were obtained.

The use of the random forest (RF) in agricultural studies has been widely observed (Jeong et al. 2016; Geetha et al. 2020; Carneiro et al. 2023) due to its robustness and generally good performance. Decision trees, and all tree-based methods such as RF, have, as some of their main advantageous features: (i) they are able to address complex nonlinear relationships without prior knowledge; (ii) they deal with heterogeneous data, whether numerical or categorical; (iii) they deal with, to a certain extent, the selection of features, leaving aside variables that do not add anything to the model; (iv) they are robust to extreme values; and (v) in the context of machine learning, they are more easily interpretable, offering less complexity to the user. The robustness of RF is based on its model ensemble characteristics. In this way, randomness is introduced into the learning process so that several different models ("weak" learners) can be obtained from a single learning set. The predictions are subsequently combined to form a robust ensemble prediction (Louppe, 2014).

2.5. Zonal application map comparison analysis

The objective in this section is to present the techniques used for comparative analysis between the resulting maps from each approach. The maps were evaluated using classification metrics (Eqs 13-15), using UF as a reference, and the performance indicated by the metrics reflects the compatibility between the UF and SUF approaches in defining classes.

$$Precision = \frac{Tp}{(Tp+Fp)} \tag{13}$$

$$Recall = \frac{Tp}{(Tp+Fn)} \tag{14}$$

$$F1 - Score = \frac{2*(Precision*Recall)}{(Precision+Recall)} \tag{15}$$

Another commonly evaluated metric is accuracy, but its interpretability can be more complex in situations of datasets with skewed distributions, as noted by Espejo-Garcia et al. (2018). The mismatch between classes on each date was also presented, thus providing a spatialized view of this incompatibility. An assessment of the conditions behind each map was performed through the analysis of plant height, vegetative indices and soil texture variability.

An economic analysis was also performed, with the objective of comparing the UF and SUF approaches using a uniform-rate (UR) scenario as a reference and considering the product consumption (CGR).

3. Results and discussion

3.1. UF: Analysis and zonal application maps

The PCA approach was primarily used in the context of UF, in which the maps were made during the agricultural season under demand and availability of field measurements. Later, the technique was also used in the SUF, resulting in maps for comparison. The main objective of the PCA was to delimit (relatively) homogeneous zones in terms of soil texture, VI and plant height variability. In Table 4 it is possible to see the main parameters of the PC analysis, previously detailed in Equations 1-9, for each field and date to create the within-season zonal application maps (Figure 4).

Table 4. Main parameters of the unsupervised framework (PC analysis) for fields and dates in the UF to create the zonal application maps.

Field	Date (mm/dd/YY)	dae	Correlation	Bartlett's test	Cum. Variance (%)	Communality
A	03/10/23	35	0.33; -0.07	8.425e-144	75	0.698; 0.699; 0.956
A	03/23/23	48	0.01; 0.06	3.974e-22	71	0.944; 0.637; 0.553
B	04/18/23	86	0.30; 0.01	3.647e-55	77	0.663; 0.658; 0.987
B	05/03/23	108	0.30; -0.15	5.708e-65	77	0.639; 0.705; 0.969
C	03/22/23	40	0.43; -0.53	0.00	85	0.829; 0.983; 0.726
C	05/11/23	88	0.22; -0.43	0.00	81	0.744; 0.993; 0.695

Obs: dae refers to days after emergence of the corresponding field data collection. Correlation refers to the correlation of crop height with ndvi and sand:clay, respectively, as the correlation matrix used as the basis for the PCA. Bartlett's test was used to determine the *p-value* and respective suitability for PCA use. The cumulative variance refers to the proportion of variability explained by the first two PCs. Communality refers to each variable's (height, ndvi, sand:clay, respectively) proportion of variability explained by the first two PCs.

As shown in Table 4, the correlation between plant height and other variables varied across fields and dates. In 3 of them, the correlation was stronger with the NDVI (although still considered weak, of approximately 0.30). For others, the correlation was greater with the soil texture and with greater intensity (between 0.40 and 0.50). There were also null correlations with one or all of the variables. This indicates that the correlation of these plant biometric characteristics varies throughout the agricultural season, as was observed during the present study. In the entire dataset with all fields and years, the general correlation (see Subsection 3.1) with VIs was relatively high (> 0.70), unlike that shown in Table 4 and this may have occurred due to the fact that in many fields there were few monitoring events, which ended up being more concentrated when there was greater correlation between VI and plant biometrics. Hoffmeister et al. (2016) evaluated the variability in detecting plant height by using multitemporal terrestrial laser scanning over 2 years and different crops. The authors used satellite images to assess relationships at different spatial resolutions, from 1-20 m. Overall, the relationship between the NDVI and plant height was poorly explained (r < 0.4) and they concluded that crop height, growth and biomass distributions were barely related to topographic and soil parameters, emphasizing the importance of using field measurements. The Bartlett test showed significance in all fields/dates of the present study at UF. This procedure was always carried out during the agricultural season for UF maps, and these results were verified for all fields/dates evaluated under this approach, indicating the adequacy of the PCA (Single, 1986). Regarding the choice of the number of PCs, this process can be performed using ad hoc rules-of-thumb, such as the evaluation of the total cumulative variance, the magnitude of the principal component variance, the scree plot or the log-eigenvalue plot. Such techniques, despite presenting a certain formality, carry the subjectivity and interpretation component and are therefore subject to the user's understanding of the analysis (Joliffe, 2002). In the present study, after exploratory evaluation of all the generated maps, the percentage of variance criterion was adopted. The choice to use the first two PCs was due to an explained variance between 70 and 90% found for all fields and dates. The communality scores (also understood as common variance), which are the proportion of each variable's variance explained by the underlying latent variables (Mittal, 2020), found in this study were all in the range of 0.65 – 0.90. Although it is just a rule-of-thumb, these values should generally be > 0.4 and ideally > 0.8 (indicating a satisfactory ability of the variables to be explained by the chosen PCs); however, in real data, they are likely between 0.4 and 0.7 (Costello and Osborne, 2019).

[Figures 4-6]
Figures 4-6. Zonal application maps of UF for fields and dates are presented in Table 4: Fields A (Figure 4), B (Figure 5), and C (Figure 6) on the first and second dates (upper and bottom rows, respectively). The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution of the present approach.

In the present study, it was observed that in general, the UF zonal application maps were visually consistent with plant height data (indicated by the points) sampled in the fields (Figures 4-6). The size of the points indicates the position and relative classification of size of the data sampled in the field for that occasion of map delineation. Once the crop height was measured and stored, it was associated with the soil texture and NDVI, and these three layers of spatially distributed data went through the UF process. With respect to the satellite images, we used the date of plant height measurements in the field as reference, and based on these, data images were obtained within an interval of up to five days (before or after) of sampling. The difficulties that weather conditions impose in the use of satellite images for precision agriculture are widely known (Shafi et al. 2019; Prudente et al. 2020), and the use of an interval of days to search for images helps to solve this problem (rainy season from December to March). The core of the UF process, based on PCA and component’s scores, ensured that each zone (indicated by classes 0, 1 and 2) had more homogeneous characteristics within itself, considering the variability of the information it carries. Thus, even in situations in which the VI does not provide relevant information (due to saturation and its low variability), the other variables will guarantee the delineation of these regions. For each region within the maps, the average plant height was presented so that the user can understand the zoning and plan the most appropriate management for each region. An important aspect of these unsupervised approach zones, is that they may vary in location in the field within a short period of time, unlike traditional management zones. This approach can be considered a necessity when focusing on dynamic factors during the agricultural season, which are strongly influenced by crop management. Méndez-Vazquez et al. (2019) noted the use of temporally dynamic zoning factors for pest management instead of only stable factors. The latter study accomplished this by adding an ecological layer that brings the dynamic component to the classic PA zoning approach. Although the design of MZs is usually based on a few years of historical VI, productivity and soil data (Plant, 2001; Nawar et al. 2017; Damian et al. 2018) there is also the possibility of delineating these zones in each agricultural season, as well as the possibility of accumulating and synthesizing several many of these maps—allowing for the improvement of understanding of the spatial response of the field to crops and the environment—a process recently facilitated by greater computational power. Few studies in the literature have evaluated and compared methods for developing the same local zonal application maps, and the studies developed by Méndez-Vázquez et al. 2019 and Bokle et al. 2023 are examples of that. Bokle et al. 2023 compared three methods of developing prescription maps for variable rates of manure application, focusing on evaluating the vulnerabilities of the farm's IT infrastructure in terms of methods. This specific scarcity of studies can be relevant when considering CGR application in cotton and may occur because VR application technologies are not widely disseminated in the CGR context (as they are for limestone and nitrogen) (Pedersen and Lind, 2017; Lowenberg-DeBoer and Erickson, 2019). The lack of such studies can also be related to the complexities of understanding the variation in the marginal benefits of inputs, the understanding of the underlying yield response function and the many interactions between environmental factors such as soil (water and nutrient content), climate, etc. (Pedersen and Lind, 2017). At the same time, many studies explore the importance of delineating zones for input management (Plant 2001; Taylor et al. 2003; Nawar et al. 2017; Ouazaa et al. 2022), what field information to use to delineate the maps (Santi et al. 2016; Ortuani et al. 2019; Mazur et al. 2022), the profitability of applications (Roberts et al., 2006; Boyer et al. 2010) and statistical frameworks (Whelan and McBratney, 2000; Buttafuoco et al. 2010).

3.2. SUF

3.2.1. Modelling data

The dataset was split into training (70%), validation (15%) and test (15%) sets. The choice to use the variables, in addition to already being supported by a previous study, also aimed to not offer complexity to the user who wishes to replicate and continue such studies.

During the modelling phase, exploratory analyses helped to understand the relationships that were evaluated. The general correlation shown in Figure 7 included all fields in which there was at least one field sampling of crop height on diverse dates during the 2.5 commercial cotton agricultural seasons. It was observed that the correlation between height and other variables widely varied during the agricultural season; e.g., on one date the correlation between height and soil texture was strong and positive, and in subsequent readings in the same plot, it could be negative or even null. The advantage of observing this type of correlation is that a greater understanding of the general dynamics of plant development is captured by remote sensing; however, attention should be given to the fact that these dynamics also vary between phenological phases (Andrea et al., 2023). In this study, strong and positive overall correlations were found for all VIs, the NIR band and days after emergence. The other bands (B, G, R) also play important associative and predictive roles (Andrea et al., 2023) in medium to strong and negative associations. Payero et al. (2004) carried out a broad comparison of VIs to estimate the height of alfalfa and grass plants (genus *Festuca*). Using a spectroradiometer to collect experimental data and linear models to evaluate the relationships, they found $R^2 > 0.90$, noting that the best results were found in the initial stages of development (more intense linearity) and problems with VI saturation occurred after plants reached a certain size. Using a field spectroradiometer, Gutierrez et al. (2012) evaluated the relationships between biomass, leaf area index (LAI) and cotton productivity with spectral indices. As in the present study, the results found were variable in relation to the phenological phase and the index or band used. Correlations of > 0.85 were found for biomass and LAI with NDVI, and even higher values were found for other indices and NIR bands (which the authors noted suffer less from saturation). Most of the recent studies relate plant phenotypic characteristics and remote sensing using unmanned aerial vehicles (UAVs) due to the greater spatial resolution of this type of image and, consequently, the quality of the information, as shown by Herr et al. 2023, in a review work on the use of UAV imagery for phenotyping cotton, corn, soybeans and wheat. However, this approach is more related to experimental conditions and production systems of reduced size and will hardly be scalable in very large commercial production areas (a few thousands of hectares would already represent and impediment to this approach).

[Figure 7]

Figure 7. Correlation between variables obtained during commercial cotton agricultural seasons and used in the predictive modelling approach.

The hyperparameter tuning process was carried out following a Bayesian optimization approach. The optimization process searched the best configuration of the hyperparameters, and the probabilistic model was updated to the following estimate of the objective function (Snoek et al. 2014). This process was performed for the following hyperparameters of the random forest regressor: number of estimators, minimum sample split, minimum sample leaf and maximum depth, while the target was represented by the mean squared error (MSE), and the values corresponding to the best model performance can be found in Table 5. More details about the hyperparameters can be found on the machine learning library page scikit-learn (https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html). The estimator performance was evaluated by k-fold cross-validation (k=5), and the Bayesian optimization process was set to 30 iterations.

Table 5. Results of the best parameter settings and targets from Bayesian optimization in hyperparameter tuning.

Num. of estimators	Min. samples split	Min. samples leaf	Max. depth	Target (MSE)
807	2	1	88	46.9

Obs: Num. of estimators is the number of trees in the forest; Min. samples split is the minimum number of samples required to split an internal node; Min. samples leaf is the minimum number of samples required to be at a lead node

3.2.2. Predictive model performance

A comparison of the predicted versus true values and the performance metrics of the regressor can be found in Figure 8 and Table 6. The results represent the performance of the model in the test set after the hyperparameters and performance metrics of the validation set were considered satisfactory.

[Figure 8]
Figure 8. Relationships between true and predicted values (crop height, cm) in the test set determined using the hyperparameter-tuned random forest regressor.

Table 6. Metrics of the test set using the hyperparameter-tuned random forest regressor.

MSE	Max. error	MAE	R^2
43.8	58.3	3.65	0.96

Obs: MSE is the mean squared error, Max. error is the maximum error, MAE is the mean absolute error and R^2 is the R-squared between the test set and the predictions of the fitted model in the test set.

The model's performance was considered satisfactory according to the metrics presented in Table 6. The maximum error is an indication of the variability present in the data, which can be explained by the number and variety of fields. The estimation of cotton height, as a single objective, is not commonly found in the literature, especially considering the use of multispectral satellite imagery, since it is more related to the use of crop height to estimate plant biomass (Kaplan et al. 2023) and yield (Herr et al. 2023). The scarcity of studies in this domain is highlighted by Weiss et al. (2020). In a meta-review on the use of remote sensing in agriculture, the authors point to plant height as being a primary variable (directly involved in the radiative transfer process), which presents low complexity in the process of modelling the relationship with radiance. The authors also noted that with regard to plant height, the literature is more developed with the use of LiDAR (light detection and ranging) and photogrammetry. In the present study, plant height estimation was a necessary step for SUF assessment, but other studies can follow these results due to the generally good performance of the predictions. Kenn et al. 2023 used orbital-level remote sensing data (Sentinel-1 and 2) to calibrate empirical models and estimate the height (also known as the leaf area index (LAI) and crop coefficient Kc) of cotton plants, finding satisfactory performances ($0.70 <= R^2 <= 0.95$) when testing several VIs. Teodoro et al. (2021) used machine learning techniques to predict agronomic variables, including plant height, in soybean crops based on images captured by sensors mounted on UAVs. The authors tested different models and variables and found the best results for plant height using the random forest and deep learning models, with $r > 0.77$. Despite the greater dissemination of images obtained through the use of UAVs, the use of orbital satellite imagery is still useful for the prediction of agronomic variables. Jamali et al. (2023) used Landsat-7 images and machine learning techniques to predict plant traits, including plant height, in wheat. Using deep learning models, they also found good results, with $R^2 = 0.82$, root mean squared error (RMSE) = 9.61, and MAE = 26.4. In cotton, Guo et al. (2022) used digital camera images to estimate plant height and biomass. Using simple linear regression models, they achieved R^2=0.83 and RMSE < 0.01. Even simple approaches such as those of these latter authors can achieve good results without requiring the use of complex models. However, it should be noted that the greatest challenges in the context of this task are after the initial "stretching" of crop development, in which the striking linearity is no longer present, the vegetation coverage is total and there is saturation of IVs, especially when occurring in large-scale commercial areas in which practically the only scalable possibility for obtaining information is through orbital-level remote sensing.

3.2.3. SUF: Unsupervised analysis and zonal application maps

The same procedure (PCA-based) was carried out in SUF after predicting plant height for each specific field and date (Table 7). The <u>correlations</u> of the estimated plant height with the NDVI and soil texture were slightly different from the patterns found in the UF. Here, higher and positive correlations of height with NDVI were found in all fields/dates (while in the UF, the values were generally intermediate), and intermediate and null correlations with soil texture (similar to the UF). This may have occurred because the remote sensing variables were predictors for plant height; therefore, this dependency relationship was already present before the unsupervised analysis. The significance of Bartlett's sphericity test also presented a different pattern of results from what was found in the UF test. In the SUF, no significant differences were found for any of the field/date combinations. These results raise important questions related to the suitability of using PCA (Jackson, 1993) in this specific framework, although some authors believe that the test might be too conservative when applied to correlation matrices.

Table 7. Main parameters of the SUF (PC analysis) for fields and dates in the UF to create the zonal application maps.

Field	Date (mm/dd/YY)	dae	Correlation	Bartlett's test	Cum. Variance (%)	Communality
A (50)	03/10/23	35	0.77; 0.06	0.14	92	0.88; 0.89; 0.99
A (50)	03/23/23	48	0.73; 0.05	0.13	90	0.87; 0.86; 0.99
B (61)	04/18/23	86	0.86; -0.08	0.17	94	0.93; 0.93; 0.99
B (61)	05/03/23	108	0.80; 0.30	0.15	93	0.90; 0.91; 0.99
C (137)	03/22/23	40	0.93; 0.58	0.20	97	0.97; 0.96; 0.99
C (13)	05/11/23	88	0.79; 0.31	0.16	93	0.90; 0.89; 0.99

Obs: Dae refers to days after emergence of the corresponding field data collection. Correlation refers to the correlation of crop height with other variables (ndvi and sand:clay) from the correlation matrix used as the basis for the PCA. Bartlett's test was used to determine the p value and respective suitability for PCA use. The cumulative variance refers to the proportion of variability explained by the first two PCs. Communality refers to each variable's (height, ndvi, sand:clay, respectively) proportion of variability explained by the first two PCs. Factorial scores refer to crop height, ndvi and sand:clay scores for each PC.

[Figures 9-11]

Figures 9-11. Zonal application maps of SUF for fields and dates as presented in Table 7: Fields A, B, and C (from top row to bottom row) on the first and second dates in each row, respectively. The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution determined via the present approach.

SUF process took more computational time than the UF for developing zonal application maps. However, the latter approach approach takes into account the availability of plant height data collected in the field, and this is unlikely to be disseminated across large areas of commercial production. These manual sampling activities, often necessary to validate the use of remote sensing technologies, have high labor costs (Tanriverdi, 2006; Ampatzidis et al. 2020) and are unlikely to be scaled up over large areas for all commercial fields with intense frequency. The agreement found in SUF within fields and dates concerning plant samples and homogeneous zones was relatively lower than the UF. Visually, one may observe regions with greater relative average plant sizes allocated to zone 0, as shown in the second date in field A. In Field B, it was also possible to observe that some of the observations collected in the field were incompatible with the map class to which they were allocated. In Field C, greater compatibility was observed between the data collected in the field and the classes defined in the SUF. An important comparative aspect between the zonal application maps of UF and SUF was that the zonal application maps of UF exhibited greater variability in their location and size. In the SUF, the zones were more permanent across field measurements and especially consistent with the soil texture. Such behaviour is directly linked to the variability of the variables used in the process. In Figures 12-14, additional details can be found about the variability and distribution of variables used in the map development process via the different approaches.

Interpretation of maps: Crop height, vegetation indices and soil texture variability

Considering each field, and the distribution of variables to aid the interpretation and comparison of the maps, in field A at UF (Figure 4), the zones differed between dates 1 and 2, while there was greater consistency between zones in SUF (Figure 9) between dates. It was also possible to observe the agreement of the zones in SUF with soil texture and plant height on date 1 and a more expressive agreement with the plant height on date 2. According to Figures 12 and 13, this field had the lowest textural and NDVI variability among the fields (except for NDVI points corresponding to regions close to the field border). Plant height (Figure 14) also had the lowest variability when compared to the other fields. It was observed for all fields a very low variability of plant height estimates coming from SUF, when compared to the values derived from UF.

[Figure 12]
Figure 12. Soil texture variability of fields A, B and C, indicated by the sand:clay ratio and used in both the UF and SUF methods to construct zonal application maps. The orange line indicates the median.

[Figure 13]

Figure 13. NDVI variability of fields A, B and C and respective dates used in the UF and SUF to construct the zonal application maps. The orange line indicates the median.

[Figure 14]

Figure 14. Crop height variability of fields A, B and C and respective dates used in the UF (plots in the left corner) and SUF (plots in the right corner) zones to construct the zonal application maps. The orange line indicates the median.

In field B, the textural variation (Figure 12) was slightly greater than that in field A (but lower than that in field C). The zonal application maps of UF (Figure 5), as it was also found in Field A, showed some variation in their location between dates, especially with regard to Class 2 (associated with the highest average plant height), and greater consistency within SUF (Figure 10). The NDVI also varied little, except for border areas and those corresponding to contour lines, which are characteristics of this field. The variability in the NDVI in this field (Figure 13) was similar to that in field A; however, in field B, the crop was at a more advanced stage of development at the first date (dae: 108) and had higher levels of NDVI, indicating that it was already in a saturation phase by the VI. The most advanced stage of plant development in field B (in dates 1 and 2) was also associated with the overall higher values of plant height (Figure 14).

In field C, it is observed the greatest soil textural variability (12), where its central region exhibits a high sand content. This led this field to present consistent patterns of plant development throughout the cycle (see maps in Figures 6 and 11), even when considering the influence of the management of different inputs (growth regulator, nitrogen fertilization and others). The strong influence of soil type (texture, physical and chemical properties), whether spatially or temporally, on vegetation growth is widely known and reported (Basso et al. 2001; Jiang et al. 2020; Feng et al. 2022). Soil texture plays an important role in determining the water content, soil nutrients and organic matter content, ultimately influencing crop development patterns through variability in vegetation dynamics. Feng et al. (2022) evaluated the variability in cotton development due to variations in soil (electric conductivity-based texture at different depths) and weather conditions. These authors showed that the association between soil features and crop growth varied widely during the season and that this correlation was stronger when crops experienced water stress and that there was a consistently strong and negative correlation between sand content and soil moisture parameters.

The effects of soil type on spatial patterns of crop growth (Basso et al. 2001) are an important concept within PA, and in this study, these effects were shown in different ways in the fields evaluated through zonal application maps. In field C, this was reflected in the NDVI and plant height variability (Figures 13 and 14) patterns during the agricultural season. Therefore, this field presented the most consistent zones between the dates and both UF and SUF approaches when compared to the other fields (Figures 6 and 11).

3.3. Zonal application map comparison analysis

Classification metrics

A confusion matrix was developed using the classes from UF as a reference (Figure 15). They do not represent an absolute truth but rather the results of an approach considered standard in this study (UF) in comparison with a with an alternative (SUF) for developing zonal application maps.

Precision is one of the simplest and most well-known metrics for evaluating classification performance. They represent a relationship between all the observations (of each class) that were labelled as positive (correctly classified in that class) and that were in fact labelled as positive (TP: true positives) and between the sum of the TP and the number of observations that were labelled as positive but that were actually negative; therefore, the TP was labelled as another class (FP: false positives). By evaluating this metric, we can understand how much we can trust the model when it predicts an observation as being of a certain class (Grandini et al. 2020). The precision for the different classes showed wide variability: values between 0.20 and 0.30 (field A), between 0.05 and 0.85 (field B), and between 0.50 and 0.80 (field C). The greatest variability was found for field B on the first date. Consistent higher values were found in field C and were generally lower in field A. Such variability in field B was lower in field A, and higher metric values were maintained in field C with the recall and, consequently, with the f1-score.

[Figure 15]

Figure 15. Metric comparison of classification reports in fields A (a), B (b) and C (c). For each field, Reports 1 and 2 refer to the first and second dates of the maps, respectively, as shown in Tables 4 and 7. Classes 0, 1 and 2 refer to the zones of application maps considering the lowest to highest crop heights, respectively.

Recall is a metric that indicates the model's ability to point to all positive observations and is therefore correctly classified in that particular class. This is done by relating the TP and the sum of the TP to observations that were labelled negative (such as other classes) but that were actually positive (of that specific class) (Grandini et al. 2020). The fact that precision and recall presented similar values is not unusual. This indicates that the algorithm classifies a similar number of false-positive and false-negative observations. Such behaviour is not necessarily bad, especially if the classes have a more balanced number of observations. Otherwise, this may indicate that the model is classifying many observations into other classes, and this should be evaluated more carefully. The F1-score is a metric that aggregates recall and precision measures and is the harmonic mean of these measures. The analysis of this metric is interesting because it helps in analysing the best trade-off between precision and recall. Grandini et al. 2020 noted that the F1-score penalizes models with greater differences between precision and recall; therefore, it is more favourable for those models that have similar values for these metrics. In the present study, the precision and recall results were relatively similar across all fields and dates, thus the F1-scores also did not differ from those of the other metrics evaluated. In general, the classification metrics point to superior compatibility between the IF and SUF maps of field C.

Spatial distribution of classification mismatch

Figure 16 summarizes the comparative analysis between UF and SUF maps by the spatial distribution of all points that showed class mismatches between the approaches, regardless of the class (0, 1 or 2). In field A (lower soil texture, NDVI and plant height variability), the lowest compatibility was found between UF and SUF when compared to all fields. The opposite results (greater variability of soil texture and plant height) and greater compatibility between the zonal application maps were found for Field C, while Field B presented intermediate results. One of the crucial points of this analysis is the layer of information on the spatial distribution of plant height and whether it is based on data collected in the field or the result of the use of a predictive model. While the sampled (and interpolated) data suggested relatively high variability, as confirmed by the field agronomists during the agricultural seasons, the results of the predictive modelling were much more conservative with respect to variability (see Figure 14).

One of the greatest difficulties at this point is that none of the layers represents the absolute truth, whether it is interpolation or prediction, but the interpolated layer has as its foundation the field truth. Another relevant point is that in fields with very intense soil texture variation, such as field C, the spatial incompatibility between classes tends to decrease. This is because such intense textural variations, represented by regions with high sand content, have a direct and consistent impact on cotton development throughout the season. However this behaviour can vary in intensity, especially due to inter annual climate variability (Guo et al. 2012). There is a need to delve deeper into these considerations related to climate and management interactions and better understand the definitions of a good zonal application map, as noted by Bökle et al. 2023, and how this might vary in different field profiles.

[Figure 16]

Figure 16. Mismatch between the classification of the zones of application between UF and SUF in fields A (a), B (b) and C (c). For each field, the first and second dates refer to the first and second dates of the maps, respectively, as shown in Tables 4 and 7. The red color corresponds to mismatches between any of the classes: 0, 1 and 2.

Cost analysis

A cost analysis is presented in Table 8, indicating how the VR approaches (UF and SUF) vary in cost compared to a uniform-dose strategy applied uniformly (UR), considering only the cost of CGR. According to this rationale, VR strategies coming from both UF and SUF had lower input costs than did UR, which were between 32 and 40% lower in all the scenarios.

Table 8. Summary of the cost analysis of applying the growth regulator in the field and dates evaluated under the uniform and variable rates techniques.

Field	Date	UR[1]		UF[1]		SUF[1]	
		Cons.[2]	Cost[3]	Cons.	Cost (%)	Cons.	Cost(%)
A	1st	25.2	151.2	18.7	111.9 (-35)	18.3	109.6 (-38)
A	2nd	25.2	151.2	18.9	113.2 (-34)	18.6	111.9 (-35)
B	1st	12.3	74	9.3	56.1 (-32)	8.7	52.2 (-42)
B	2nd	12.3	74	9.2	55.0 (-35)	8.6	51.8 (-43)
C	1st	27.5	164.9	19.8	118.9 (-39)	19.8	1178.5 (-39)
C	2nd	27.5	164.9	20.1	120.6 (-37)	19.6	117.5 (-40)

[1] UR refers to the uniform rate of application, while UF and SUF refer to the variable rate approaches assessed in this study. For simplicity, a rate of 0.16 L ha^{-1} for the UR and rates of 0.0, 0.15 and 0.20 L ha^{-1} for classes 0, 1 and 2 in both the UF and SUF, respectively, were used as references for CGR operational consumption. These values were based on the actual consumption in some of the fields during the 2022-2023 agricultural season. [2] The consumption of input (L) was calculated following the following formula: UR: $Cons. = (\Sigma areasofclasses) * UR$; UF/SUF: $Cons. = \sum_{class=0}^{2}(Area * Rateofclass)$, where the area of each class was calculated based on the map of application and the respective spatial resolution. For UF/SUF, the consumption presented is the sum of all the classes, resulting in the total consumption of the fields. [3] The cost of the applied input in the field ($) was obtained by the association of consumption with the average cost of CGR. Mepiquat chloride was the CGR considered for this analysis, at the price of 6 $ L^{-1}.

Some important points should be considered in this type of analysis: (i) In practice, the application cost for all scenarios will be higher because it will include costs related to machinery, labor, and other factors beyond the input. VR applications are exclusive ones, i.e., a mechanized operation must perform only the application of CGR; while in UR, the CGR is applied at a single rate together with other inputs, such as insecticides, decreasing the application cost of this strategy. (ii) For simplification purposes, representative doses of applications of these fields and farms were used, but they vary between applications within the same field. There might be situations in which the field agronomist notes the need for an application rate even in the smallest class (class 0), which can result in VR strategy costs similar to those of the UR strategy; (iii) moreover, there should not always exist the necessity or the (machinery) availability for CGR application in VR. According to the field data evaluated in this study, around 10 CGR applications are usually taken during a cropping season on those fields. There is no availability of machinery suitable for VR applications for all fields and inputs, which is a reality in very large production areas, and most applications end up following the UR strategy (CGR applied together with other inputs). In addition, there are fields where low (relative) textural variability, and consequently plant development, can lead to a natural choice for the UR application strategy. Therefore, for VR applications to be more profitable in these scenarios, other operational costs must not exceed the margin of savings they provide from the applied input. As noted by Lawes and Robertston (2011), VR technology can add value to many fields, especially those with greater heterogeneity; however, gains in this type of management are usually small, and the technology must be implemented cheaply to become viable. Although all the fields presented similar values for reducing input costs under VR approaches in this study, the results suggest agreement with the general knowledge that VR techniques in site-specific management usually bring potentially greater benefits in fields with greater heterogeneity concerning soil types (Pannell et al 2019; Späti et al. 2021). Considering that most of the studies evaluating the impacts of using VR are focused on nutrient management (economic and/or environmental benefits), a large number of studies indicate that the profitability of VR techniques must be evaluated in relation to the local context (Plant, 2001; Lawes and Robertson 2011; Fabiani et al. 2020; Tenreiro et al. 2023); studies that found no significant differences from a uniform rate approach (Boyer et al. 2011); and studies that noted that the overall economic benefits of this type of management in small-scale farming are low (Späti et al. 2021).

3.4. Future steps

Future studies may consider a more detailed investigation on the aspects of crop management, climate conditions, their interactions and how they impact on crop height evolution during the agricultural season. The use of yield maps can also be considered in a results analysis. Other sampling methods and arrangements also might be relevant to improve the spatial distribution layer of crop height as input.

4. Conclusions

The present study used a fairly vast dataset of field-measured cotton heights from hundreds of fields under commercial conditions to perform a comparative evaluation of two approaches to obtain zonal application maps. The general methodology for both approaches consisted of a PCA-based framework based on crop height, vegetation indices and soil texture. The approaches differed in the way they obtained plant height data: field measured based or based on model predictions. The approach based on field data collected was used as a reference for capturing local variability and building zonal application maps to support the application of variable rate growth regulators in large commercial cotton fields. The use of field data provided a more accurate view of the plant height variability found in the field, resulting in maps that address each reading and, consequently, the map in a unique and specific way. Through the machine learning model-based approach, we also showed that there is potential for the use of predictive modelling to assist in the construction of zonal application maps but that the adherence to real patterns of crop growth variability found in the field is highly variable and must be assessed on a case-by-case basis. When the soil texture variability was more intense, since crop development in the field followed a unique and consistent pattern throughout the entire cycle, the maps were more consistent between the readings and approaches. The lower the textural variability of a field is, the lower is the occurrence of consistent patterns of growth and consequently of zones in the maps between readings and the lower the consistency between approaches. Although the use of inputs in VR strategies is usually more efficient, issues related to the availability of machinery and the cost of dedicated operation must be included in planning and evaluation. continues to be a reality for large properties due to the combined effects of the availability of machines for VR applications and the existence of fields with less (relative) textural and plant development variability.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to their origin in a private company, but may be available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Andrea MCS, Oliveira CF designed the experiment; Andrea MCS, Santos RC, Rodrigues Junior EF, Bianchi LM, Gouveia CM analysed the data; Andrea MCS led the writing of the manuscript, Oliveira CF, Mota FCM, Oliveira RS, Barbosa VGS, Silva MAB revised the work; All authors read and approved the final manuscript.

Acknowledgements
The authors are grateful to the agronomists and technicians of Amaggi for the field monitoring and data collection activities, as also for the useful suggestions made during the conduct of the experiments.

References

Alesso CA, Martin NF. Spatial and temporal variability of corn response to nitrogen and seed rates. Agronomy Journal. 2023; 1-20. doi: 10.1002/agj2.21471.

Ampatzidis Y, Partel V, Costa L. Agroview: Cloud-based application to process, analyze and visualize UAV- collected data for precision agriculture applications utilizing artificial intelligence. Computers and Electronics in Agriculture. 2020; 174: 105457. doi: 10.1016/j.compag.2020.105457.

Andrea MCS da, Nascimento JPF de O, Mota FCM, Oliveira R de S. Predictive framework of plant height in commercial cotton fields using a remote sensing and machine learning approach. Smart Agricultural Technology. 2023; 4: 100154. doi: 10.1016/j.atech.2022.100154.

Basso B, Ritchie JT, Pierce FJ, Braga RP, Jones JW. Spatial validation of crop models for precision agriculture. Agricultural Systems. 2001; 68: 97-112.

Berger PG, Lima TC, Oliveira R. Algodão no cerrado: logística e operações práticas: Volume 1: Do planejamento agrícola à aplicação de reguladores de crescimento. Viçosa, MG: UFV, CEAD. 2019; https://serieconhecimento.cead.ufv.br/edicoes/algodao-no-cerrado-logistica-e-operacoes-praticas-volume-1/. Accessed Nov 7 2023.

Bökle S, Karampoiki M, Paraforos DS, Griepentrog HW. Using an open source and resilient technology framework to generate and execute prescription maps for site-specific manure application. Smart Agricultural Technology. 2023; 5:100272. doi: 10.1016/j.atech.2023.100272.

Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. Interdiscip. J. Inf. Knowl. Manag. 2019; 14:45–79. doi: 10.48550/arXiv.1809.03006.

Boyer CN, Brorsen BW, Solie JB, Raun WR. Profitability of variable rate nitrogen application in wheat production. Precision Agric. 2011; 12: 473-487. doi: 10.1007/s11119-010-9190-5.

Buttafuoco G, Castrignanò A, Colecchia AS, Ricca N. Delineation of Management Zones Using Soil Properties and a Multivariate Geostatistical Approach. Ital. J. Agron. 2010; 4: 323-332.

Carneiro FM, Brito Filho AL de, Ferreira FM, Seben Junior GF, Brandão ZN, Silva RP, Shiratsuchi LS. Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. Smart Agricultural Technology. 2023; 100292. doi:10.1016/j.atech.2023.100292.

CONAB Companhia Nacional de Abastecimento - Séries Históricas. 2023; https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras/itemlist/category/898-algodao. Accessed 15 Dec 2023.

Cordoba MA, Bruno CI, Costa JL, Peralta NR, Balzarini MG.
Protocol for multivariate homogeneous zone delineation in precision agriculture. Biosystems Engineering. 2016; 143: 95-107. doi: 10.1016/j.biosystemseng.2015.12.008.

Costello AB, Osborne, J. Best practices in exploratory factor analysis: four
recommendations for getting the most from your analysis. Practical Assessment, Research, and Evaluation. 2019; 10(7). doi: 10.7275/jyj1-4868.

Damian JM, Pias OHC, Cherubin MR, Fonseca AZ, Fornari EZ, Santi AL. Applying the NDVI from satellite images in delimiting management zones for annual
crops. Scientia Agricola. 2018; 77(1). doi:10.1590/1678-992X-2018-0055.

Doerge TA. Management zones concepts Site-specific management guidelines. Norcross: Potash & Phosphate Institute. 2000; 135 p. http://www.ipni.net/publication/ssmg.nsf/0/C0D052F04A53E0BF852579E500761AE3/$FILE/SSMG-02.pdf. Accessed in 20 Nov 2023.

Echer FR, Rosolem CA. Reguladores De crescimento: Da Fisiologia à Aplicação
In: Fisiologia aplicada ao manejo do algodoeiro Instituto Mato-Grossense do Algodão. 2022; https://imamt.org.br/wp-content/uploads/2022/02/boletim_fisiologia_2022_VF_WEB.pdf. Accessed 10 Nov 2023.

Espejo-Garcia B, Martinez-Guanter J, Pérez-Ruiz M, Lopez-Pellicer FJ, Zarazaga-Soria J. Machine learning for automatic rule classification of agricultural T regulations: A case study in Spain. Computers and Electronics in Agriculture. 2018; 150:343-352. doi: 10.1016/j.compag.2018.05.007.

Fabiani S, Vanino S, Napoli R, Zajíček A, Duffková R, Evangelou E, Nino P. Assessment of the economic and environmental sustainability of Variable Rate Technology (VRT) application in different wheat intensive European agricultural areas. A Water energy food nexus approach. Environmental Science and Policy. 2020; 114: 366-376.

Feng A, Zhou J, Vories ED, Sudduth KA. Quantifying the effects of soil texture and weather on cotton development and yield using UAV imagery. Precision Agriculture. 2022; 23: 1248-1275. doi: 10.1007/s11119-022-09883-6.

Fridgen JJ, Kitchen NR, Sudduth KA, Drummond ST, Wiebold WJ, Fraisse CW. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. Agron. J. 2004; 96: 100-108.

Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S, Janani AP. An Effective Crop Prediction Using Random Forest Algorithm. International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India. 2020; 1-5, doi: 10.1109/ICSCAN49426.2020.9262311.

Gitelson AA. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Remote Sens. Environ. 2004; doi:10.1078/0176-1617-01176.

Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 1996; 58: 289–298.

Grandini M, Bagli E, Visani G. Metrics for Multi-Class Classification: an Overview. ArXiv. 2020; 2008.05756.

Guo C, Liu L, Zhang K, Sun H, Zhang Y, Li A, Bai Z, Dong H, Li C. High-throughput estimation of plant height and above-ground biomass of cotton using digital image analysis and Canopeo. Technology in Agronomy. 2022; 2:4. doi: 10.48130/TIA-2022-0004.

Gutierrez M, Norton R, Thorp KR, Wang G. Association of spectral reflectance indices with plant growth and lint yield in upland cotton. Crop Sci. 2012; 161(2): 165-173. doi:10.2135/cropsci2011.04.02227

Gutman G, Skakun S, Gitelson A. Revisiting the use of red and near-infrared reflectances in vegetation studies and numerical climate models. Science of Remote Sensing. 2021; 4:100025. doi: 10.1016/j.srs.2021.100025.

Herr AW, Adak A, Carroll ME, Elango D, Kar S, Li C, Jones Se, Carter AH, Murray SC, Paterson A, Sankaran S, Singh A, Singh AK. Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. Crop Science. 2023; 63:1722–1749. doi: 10.1002/csc2.21028.

Hoffmeister D, Waldhoff G, Korres W, Curdt C, Bareth G. Crop height variability detection in a single field by multi-temporal terrestrial laser scanning. Precision Agric. 2016; 17:296-312.

Jackson DA. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology. 1993; 74(8): 2204-2214.

Jamali M, Bakhshandesh E, Yeganeh B, Özdoğan M. Development of machine learning models for estimating wheat biophysical variables using satellite-based vegetation indices. Advances in Space Research. 2023; 73(1): 498-513. doi: 10.1016/j.asr.2023.10.004.

Janse PV, Deshmukh RR, Randive PU. Vegetation indices for crop management: a review. IJRAR. 2018; 6(1): 413:415.

Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, et al. Random Forests for Global and Regional Crop Yield Predictions. 2016; PLOS ONE 11(6): e0156571 doi: 10.1371/journal.pone.0156571

Jiang P, Ding W, Yuan Y, Ye W. Diverse response of vegetation growth to multi-time-scale drought under different soil textures in China's pastoral areas. Journal of Environmental Management. 2020; 274: 110992. doi: 10.1016/j.jenvman.2020.110992.

Jolliffe IT. Principal component analysis, 2nd edn. New York, NY: Springer-Verlag; 2002.

Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A. 2016; 374: 20150202. doi: 10.1098/rsta.2015.0202.

Kaplan G, Fine L, Lukyanov V, Malachy N, Tanny J, Rozenstein O. Using Sentinel-1 and Sentinel-2 imagery for estimating cotton crop coefficient, height, and Leaf Area Index. Agricultural Water Management. 2023; 276: 108056. doi: 10.1016/j.agwat.2022.108056.

Keller W, Borkowski A. Thin plate spline interpolation. J Geod. 2019; 93: 1251–1269. doi: 10.1007/s00190-019-01240-2.

Lawes RA, Robertson MJ. Whole farm implications on the application of variable rate technology to every cropped field. Field Crops Research. 2011; 124: 142-148. doi: 10.1016/j.fcr.2011.01.002.

Liu L-W, Ma X, Wang Y-M, Lu C-T, Lin W-S. Using artificial intelligence algorithms to predict rice (Oryza sativa L.) growth rate for precision agriculture. Computers and Electronics in Agriculture. 2021; 187: 106286. doi: 10.1016/j.compag.2021.106286.

Louppe G. Understanding Random Forests From Theory to Practice. [Doctoral dissertation, University of Liège]. 2014. doi: 10.48550/arXiv.1407.7502.

Lowenberg-DeBoer J, Erickson B. Setting the Record Straight on Precision Agriculture Adoption. Agronomy Journal. 2019; 111(4): 1-18. doi: 10.2134/agronj2018.12.0779.

Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Mazur P, Gozdowski D, Wójcik-Gront E. Soil Electrical Conductivity and Satellite-Derived Vegetation Indices for Evaluation of Phosphorus, Potassium and Magnesium Content, pH, and Delineation of Within-Field Management Zones. Agriculture. 2022; 12:883. doi: 10.3390/agriculture12060883.

Méndez-Vázquez LJ. Lira-Noriega A, Lasa-Covarrubias R, Cerdeira-Estrada S. Delineation of site-specific management zones for pest control purposes: Exploring precision agriculture and species distribution modeling approaches. Computers and Electronics in Agriculture. 2019; 167: 105101. doi: 10.1016/j.compag.2019.105101.

Mittal P. A multicriterion decision analysis based on PCA for analysing the digital technology skills in the effectiveness of government services. International Conference on Decision Aid Sciences and Application (DASA). 2020; 490-494. doi: 10.1109/DASA51403.2020.9317241

Mro´z M, Sobieraj A. Comparison of several vegetation indices calculated on the basis of a seasonal spot XS time series, and their suitability for land cover and agricultural crop identification, Tech. Sci. 2004; 7.

Narasimhulu Y, Suthar A, Pasunuri R, Venkaiah VC. CKD-Tree: An Improved KD-Tree Construction Algorithm. CEUR Workshop Proc. 2021, 2786, 211–218.

Nawar S, Corstanje R, Halcro G, Mulla D, Mouazen AM. Delineation of Soil Management Zones for Variable-Rate Fertilization. Adv. Agron. 2017; 143: 175-245. doi: 10.1016/bs.agron.2017.01.003.

Nogueira F. (2014) Bayesian Optimization: Open source constrained global optimization tool for Python. 2014. https://github.com/fmfn/BayesianOptimization. Accessed in 15 Oct 2023.

OECD-FAO. 'OECD-FAO Agricultural Outlook 2023-2032', Technical report, OECD, and Food and Agriculture Organization of the United Nations. https://www.oecd-ilibrary.org/sites/f1e4942f-en/index.html?itemId=/content/component/f1e4942f-en. 2023. Accessed 20 Dec 2023.

Ortuani B, Sona G, Ronchetti G, Mayer A, Facchi A. Integrating Geophysical and Multispectral Data to Delineate Homogeneous Management Zones within a Vineyard in Northern Italy. Sensors. 2019; 19: 3974. doi:10.3390/s19183974.

Ouazaa S, Jaramillo-Barrios CI, Chaali N, Amaya YMQ, Carvajal JEC, Ramos OM. Towards site specific management zones delineation in rotational cropping system: Application of multivariate spatial clustering model based on soil properties. Geoderma Regional. 2022; 30: e00564. doi: 10.1016/j.geodrs.2022.e00564.

Pannell D, Gandorfer M, Weersink A. How flat is flat? Measuring payoff functions and the implications for site-specific crop management. Computers and Electronics in Agriculture. 2019; 162: 459-465. doi: 10.1016/j.compag.2019.04.011.

Payero JO, Neale CMU, Wright JL. Comparison of eleven vegetation indices for estimating plant height of Alfafa and Grass. Appl. Eng. Agric. 2004; 20(3): 385-393.

Pedersen SM, Lind KM. Precision Agriculture – From Mapping to Site-Specific Application. In: Pedersen SM, Lind KM, editors. Precision Agriculture: Technology and Economic Perspectives, Progress in Precision Agriculture. Frederiksberg: Springer International Publishing AG; 2017. p.1-20.

Pierce FJ, Nowak P. Aspects of precision agriculture. Advances in Agronomy. 1999; 67: 1-85.

Planet. Planet imagery product specifications. Available at https://assets.pl anet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf. 2022; Access on 10 Jun 2023.

Plant RE. Site-specific management: the application of information technology to crop production. Computers and Electronics in Agriculture. 2001; 9:29-30.

Prudente VHR, Martins VS, Vieira DC, Silva NRF, Adami M, Sanches IDA. Limitations of cloud cover for optical remote sensing of agricultural areas across South America. Remote Sensing Applications: Society and Environment. 2020; 20: 100414. doi: 10.1016/j.rsase.2020.100414.

Roberts RK, English BC, Larson JA. The Variable-Rate Input Application Decision for Multiple Inputs with Interactions. Journal of Agricultural and Resource Economics. 2006; 31(2): 391-413.

Santi AL, Damian JM, Cherubin MR, Amado TJC, Eitelwein MT, Vian AL, Herrera WFB. Soil physical and hydraulic changes in different yielding zones under no-tillage in Brazil. Afr. J. Agric. Res. 2016; 11(15): 1326-1335. doi: 10.5897/AJAR2015-10643.

Sastry SP, Zalaa V, Kirbya RM.Thin-Plate-Spline Curvilinear Meshing on a Calculus-of-Variations Framework. Procedia Engineering. 2015; 124:135-147. doi: 10.1016/j.proeng.2015.10.128.

Shafi U, Mumtaz R, García-Nieto J, Hassan SA, Zaidi SAR, Iqbal N. Precision Agriculture Techniques and Practices: From Considerations to Applications. Sensors. 2019; 19: 3796. doi: 10.3390/s19173796.

Silvertooth JC, Wrona AF, Guthrie DS, Hake K, Kerby T, Landivar JA. Vigor indices for cotton management. Cotton Physiology Today: Newsletter of the Cotton Physiology Education Program - National Cotton Council. 1996; v.7, n.3.

Single JR. A priori estimation of sample size and number of variables for principal component analysis. Bulletin of the Southern California Academy of Sciences. 1986; 85(2): 123–125. doi: 10.3160/0038-3872-85.2.123.

Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems 25 (NIPS 2012). 2012. https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf. Accessed in 15 Nov 2023.

Souza AP de, Mota LL da, Zamadei T, Martim CC, Almeida FT de, Paulino J. Classificação climática e balanço hídrico climatológico do estado de Mato Grosso. Nativa. 2013; 34:43-1.

Späti K, Huber R, Finger R. Benefits of Increasing Information Accuracy in Variable Rate Technologies. Ecological Economics. 2021; 185: 107047. doi: 10.1016/j.ecolecon.2021.107047.

Taylor R, Fulton J. Sensor–based variable rate application for cotton. Oklahoma Cooperative Extension Service. Oklahoma State University. Stillwater, Oklahoma. 2010; 1-8.

Taylor JC, Wood GA, Earl R, Godwin RJ. Soil Factors and their Influence on Within-field Crop Variability, Part II: Spatial Analysis and Determination of Management Zones. Biosystems Engineering. 2003; 84(4): 441-453. doi: 10.1016/S1537-5110(03)00005-9.

Tanriverdi, C. A review of remote sensing and vegetation indices in precision farming. J. Sci. Eng. 2006; 9: 69–76.

Teodoro PE, Teodoro LPR, Baio FHR, Silva Junior CA, Santos RG, Ramos APM, Pinheiro MMF, Osco LP, Gonçalves WN, Carneiro AM, Marcato Junior J, Pistori H, Shiratsuchi LS. Predicting Days to Maturity, Plant Height, and Grain Yield in Soybean: A Machine and Deep Learning Approach Using Multispectral Data. Remote Sensing. 2021; 13: 4632. doi: 10.3390/rs13224632.

Tenreiro TR, Avillez F, Gómez JA, Penteado M, Coelho JC, Fereres E. Opportunities for variable rate application of nitrogen under spatial water variations in rainfed wheat systems—an economic analysis. Precision Agriculture. 2023; 24: 853-878. doi: 10.1007/s11119-022-09977-1.

Tucker CJ. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sensing of Environment. 1979; 8:127-150.

Vaz CMP, Franchini JC, Speranza EA, Inamasu RY, Jorge LA de C, Rabello LM, Lopes I de ON, Chagas S das, Souza JLR de, Souza M de, Pires A, Schepers J. Zonal Application of Plant Growth Regulator in Cotton to Reduce Variability and Increase Yield in a Highly Variable Field. Agronomy and Soils. 2023; 27: 60-73.

Verrelst J, Malenovsky Z, Van der Tol C, Camps-Valls G, Gastellu-Etchegorry J-P, Lewis P, North P, Moreno J. Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. Surv. Geophys. 2019; 40: 589–629, doi: 10.1007/s10712-018-9478-y.

Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment. 2020; 236: 111402. doi: 10.1016/j.rse.2019.111402.

Whelan BM, McBratney AB. The ‘‘Null Hypothesis’’ of Precision Agriculture Management. Precision Agriculture. 2000; 2: 265-279.

Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. Journal of Electronic Science and Technology. 2019; 17(1) doi: 10.11989/JEST.1674-862X.80904120.

Xue J, Su B. Significant Remote Sensing Vegetation Indices: A Review of Development and Applications. Hindawi Journal of Sensors. 2017; 1353691. doi: 10.1155/2017/1353691.

Figures

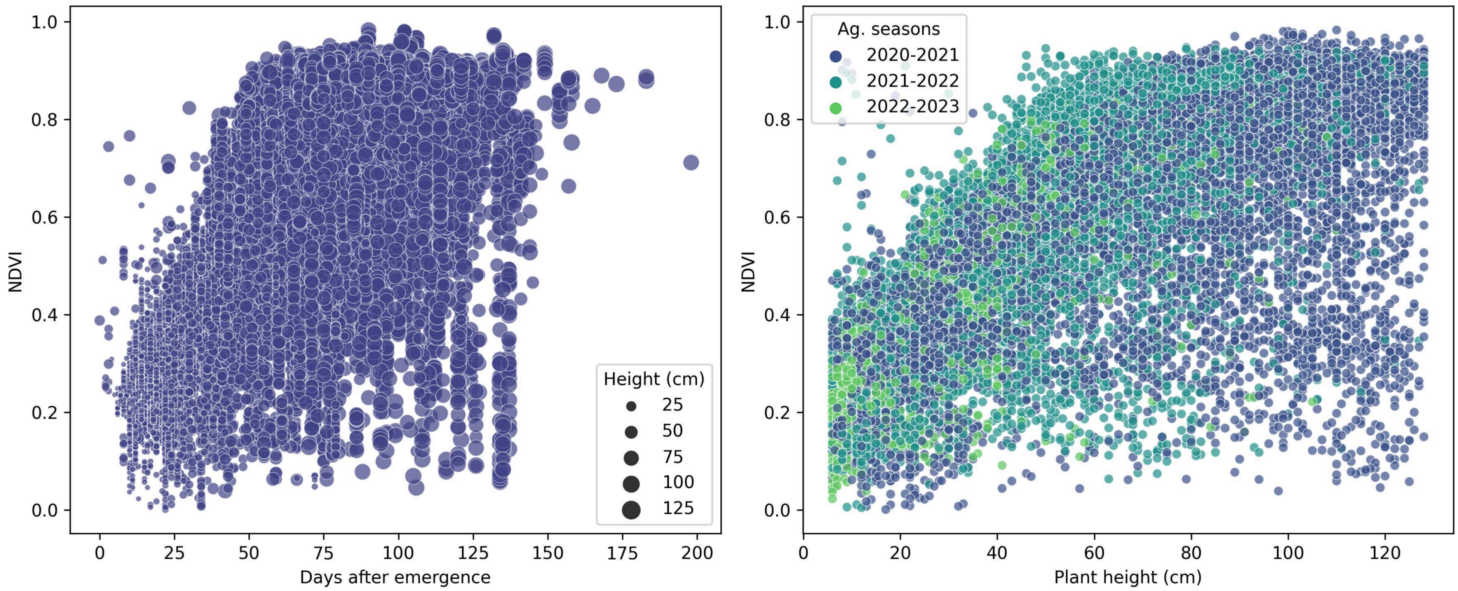

Figure 1

Relationships between days after emergence and the NDVI of cotton fields (left figure) and between the measured cotton height and the NDVI across agricultural seasons (right figure) of the fields assessed in the present study.

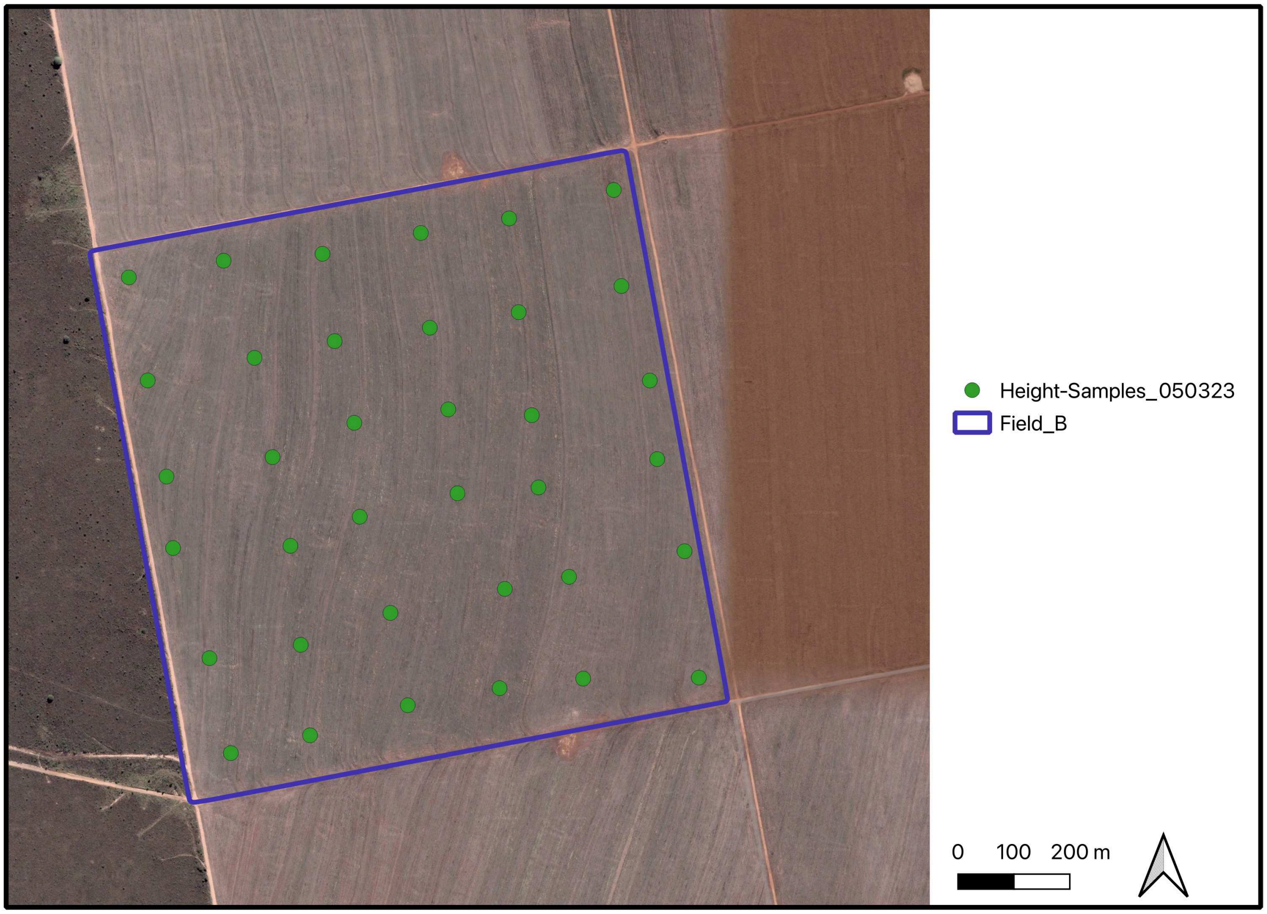

Figure 2

Distribution of plant height measurement points in Field A during the 2022-2023 agricultural season.

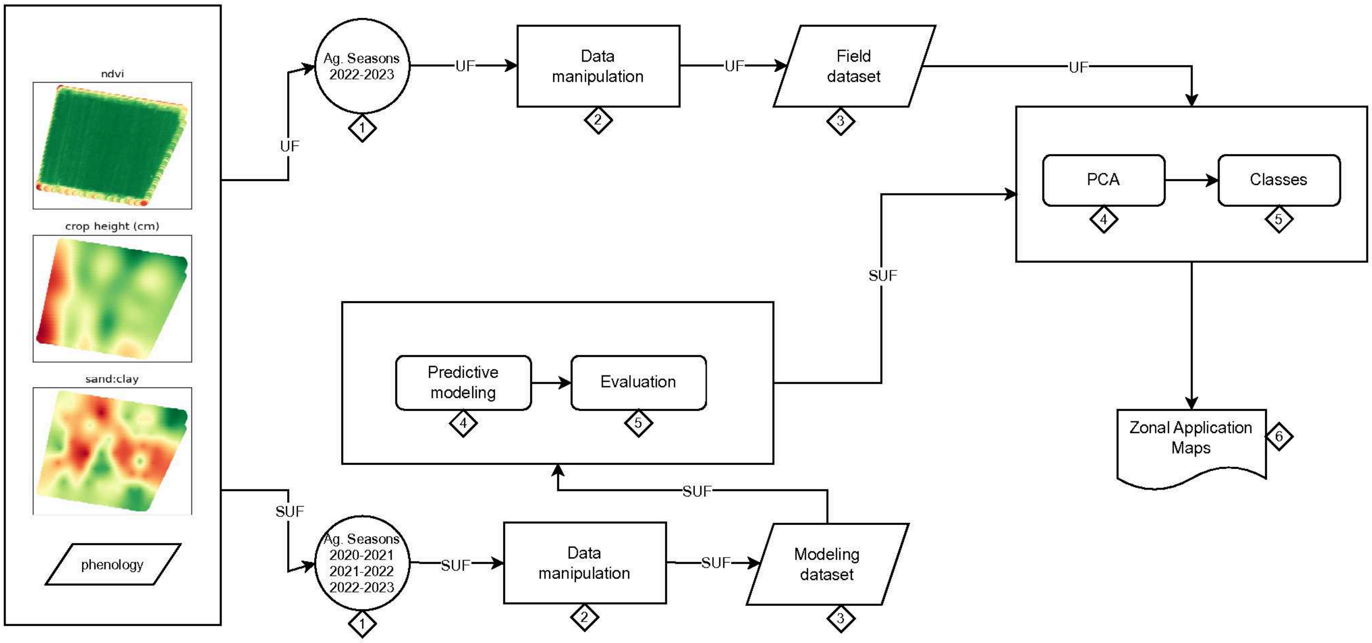

Figure 3

Flowchart summarizing the methodological approaches used in this study.

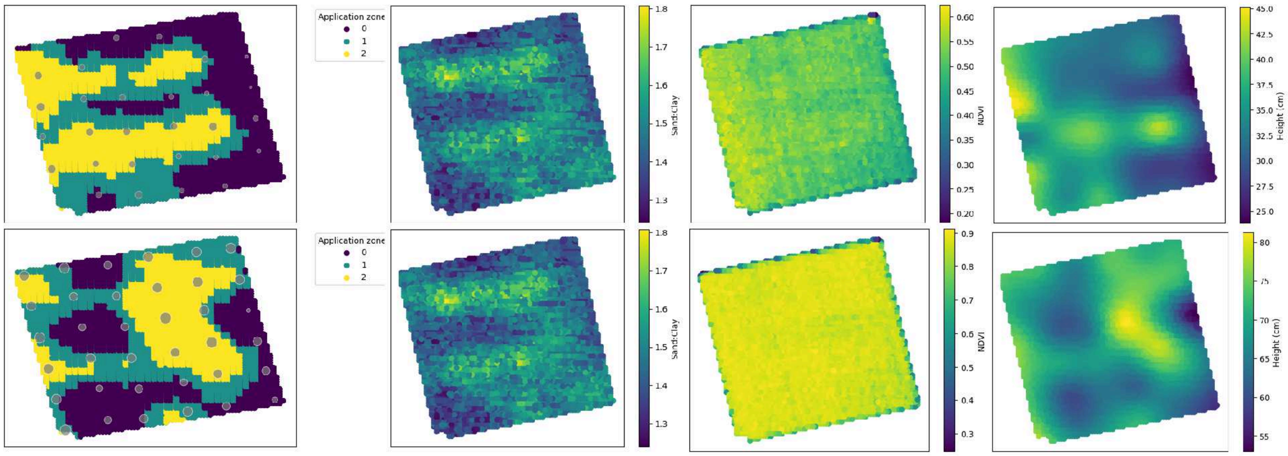

Figure 4

Zonal application maps of UF for fields and dates are presented in Table 4: Fields A (Figure 4), B (Figure 5), and C (Figure 6) on the first and second dates (upper and bottom rows, respectively). The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution of the present approach.

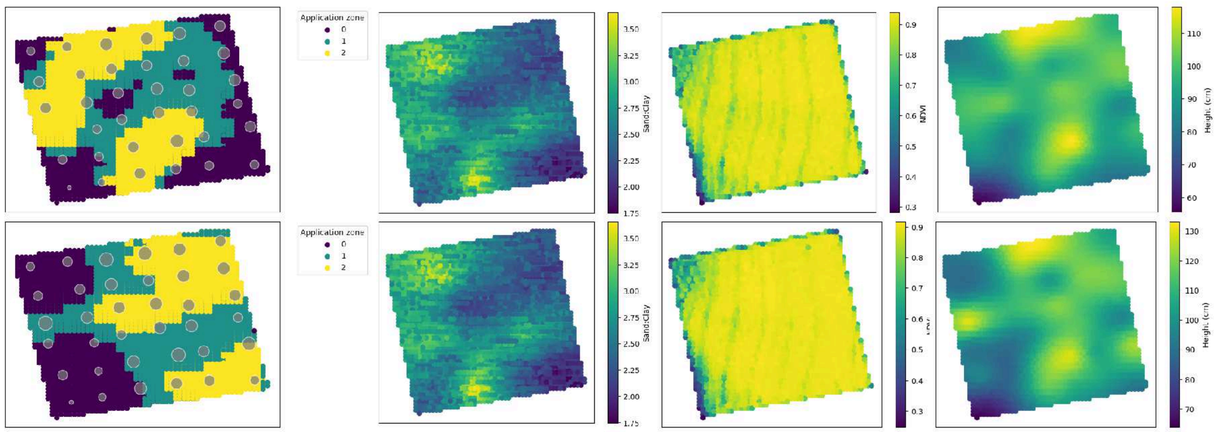

Figure 5

Zonal application maps of UF for fields and dates are presented in Table 4: Fields A (Figure 4), B (Figure 5), and C (Figure 6) on the first and second dates (upper and bottom rows, respectively). The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution of the present approach.

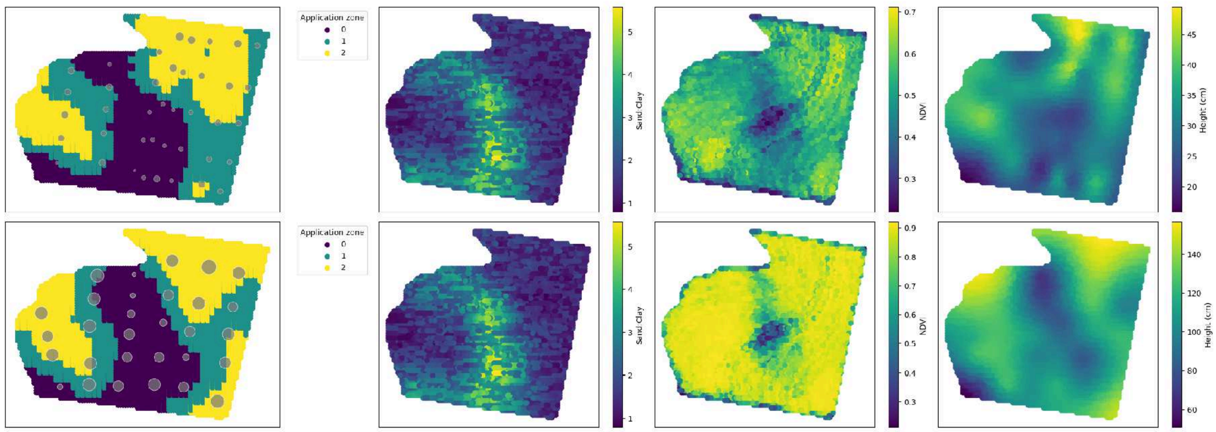

Figure 6

Zonal application maps of UF for fields and dates are presented in Table 4: Fields A (Figure 4), B (Figure 5), and C (Figure 6) on the first and second dates (upper and bottom rows, respectively). The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution of the present approach.

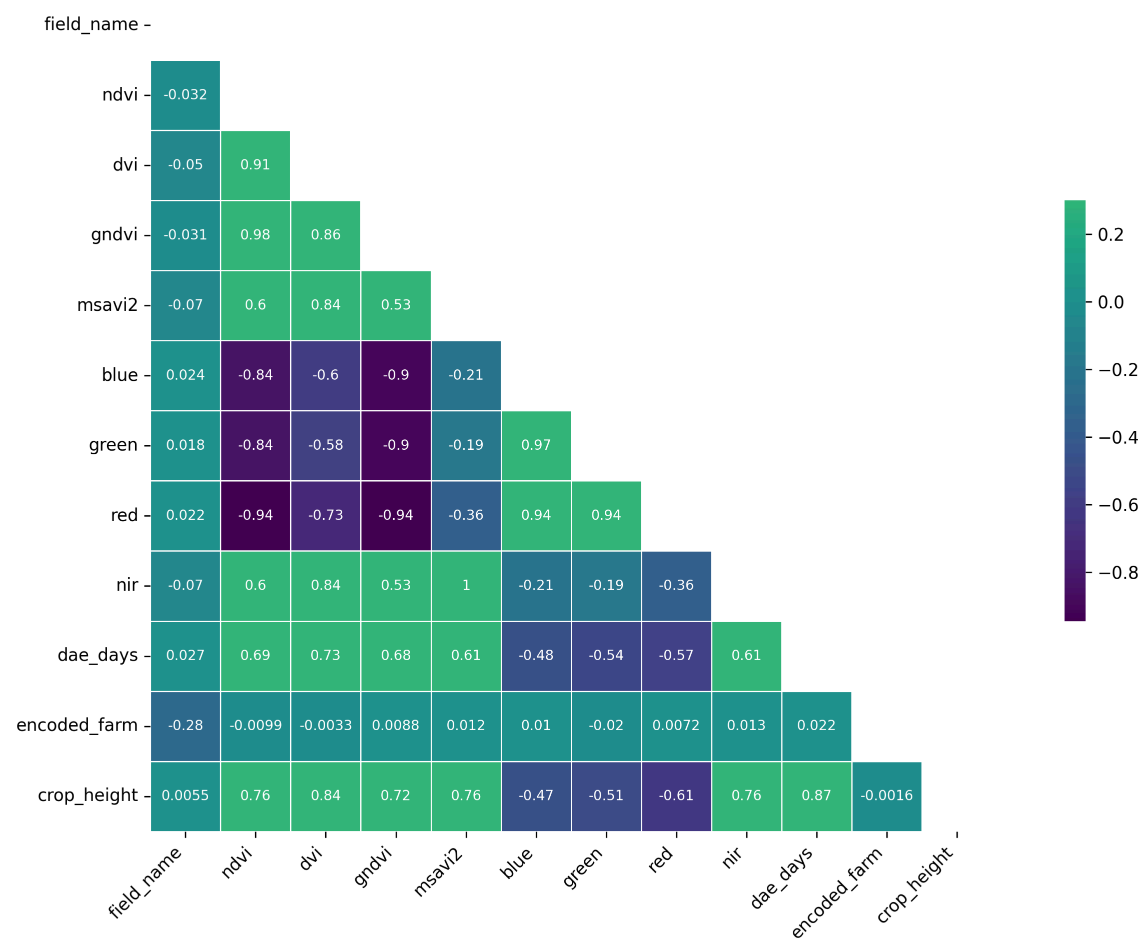

Figure 7

Correlation between variables obtained during commercial cotton agricultural seasons and used in the predictive modelling approach.

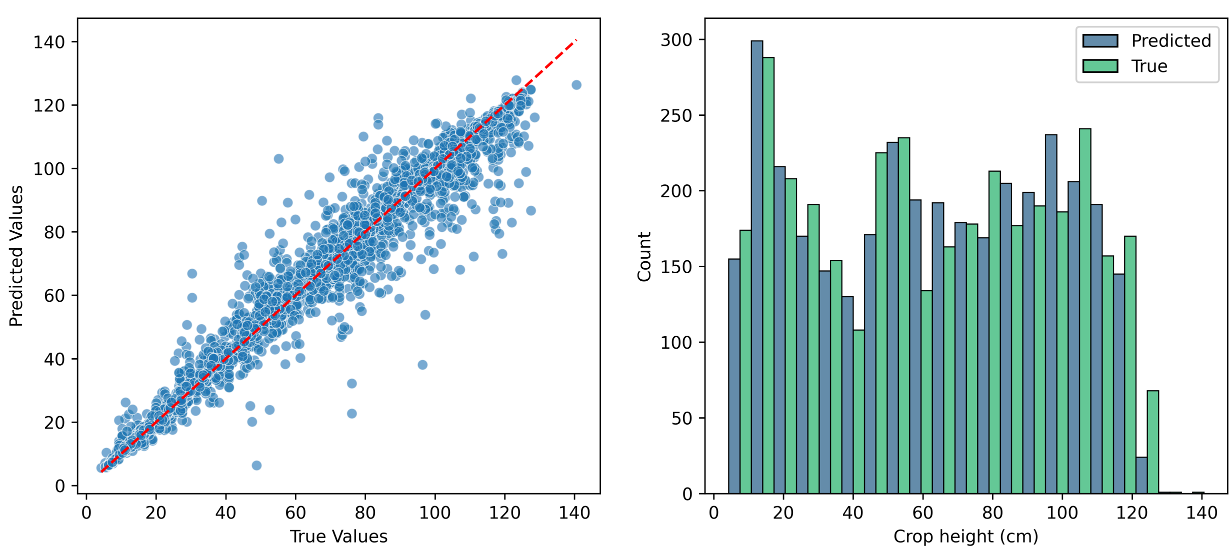

Figure 8

Relationships between true and predicted values (crop height, cm) in the test set determined using the hyperparameter-tuned random forest regressor.

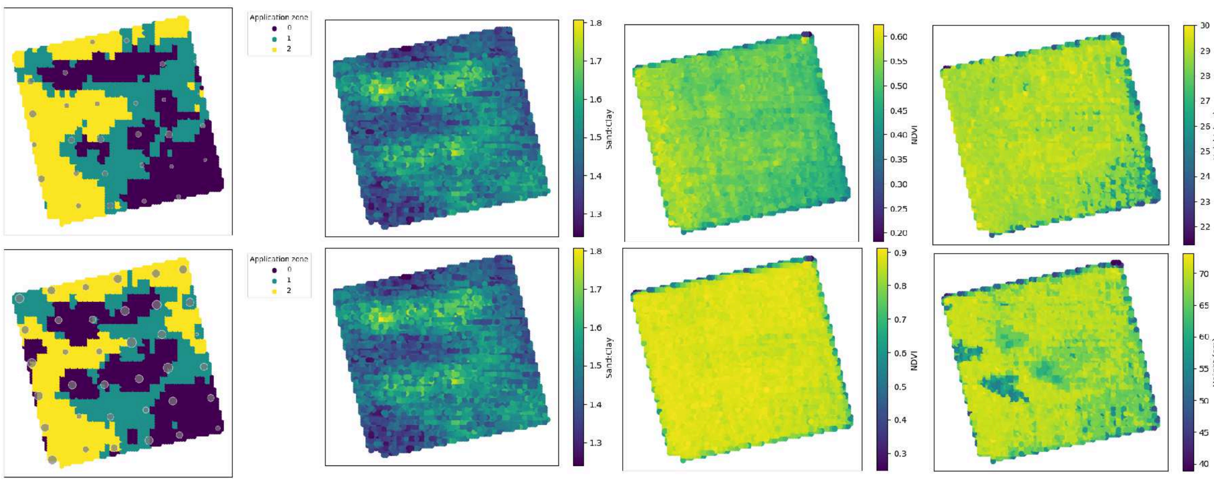

Figure 9

Zonal application maps of SUF for fields and dates as presented in Table 7: Fields A, B, and C (from top row to bottom row) on the first and second dates in each row, respectively. The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution determined via the present approach.

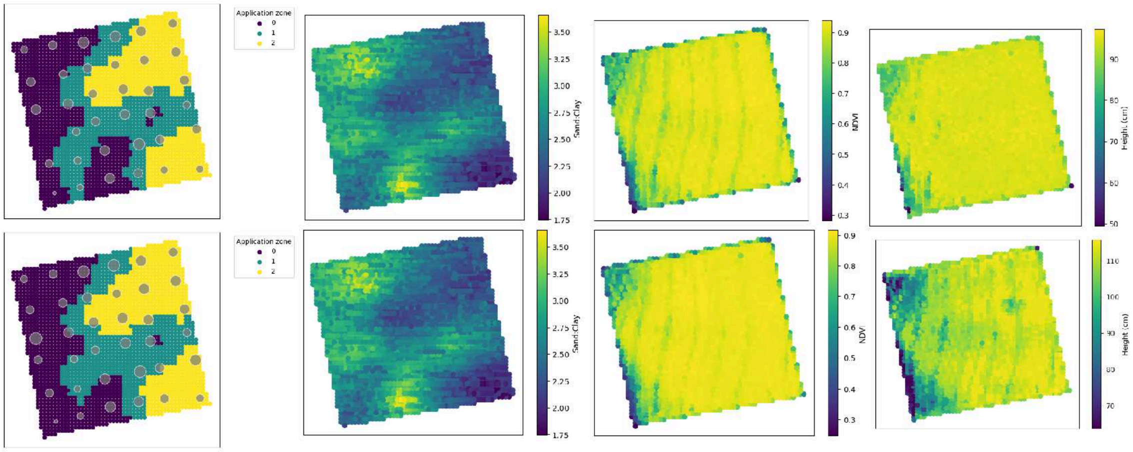

Figure 10

Zonal application maps of SUF for fields and dates as presented in Table 7: Fields A, B, and C (from top row to bottom row) on the first and second dates in each row, respectively. The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution determined via the present approach.

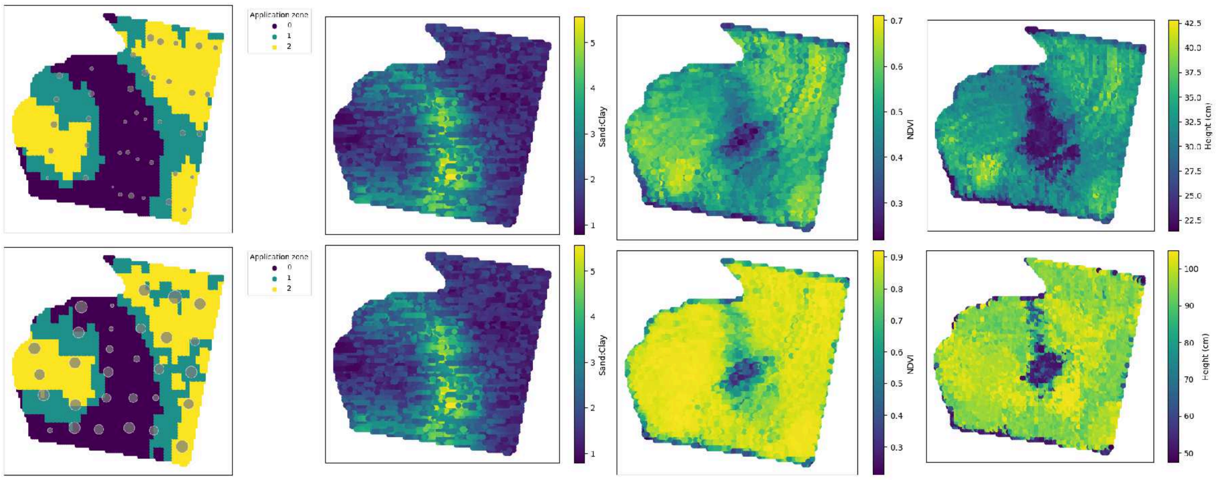

Figure 11

Zonal application maps of SUF for fields and dates as presented in Table 7: Fields A, B, and C (from top row to bottom row) on the first and second dates in each row, respectively. The first plot in the left corner is the zonal application map, the second plot is the soil texture distribution, the third plot is the NDVI distribution, and the plot in the right corner corresponds to the crop height distribution determined via the present approach.

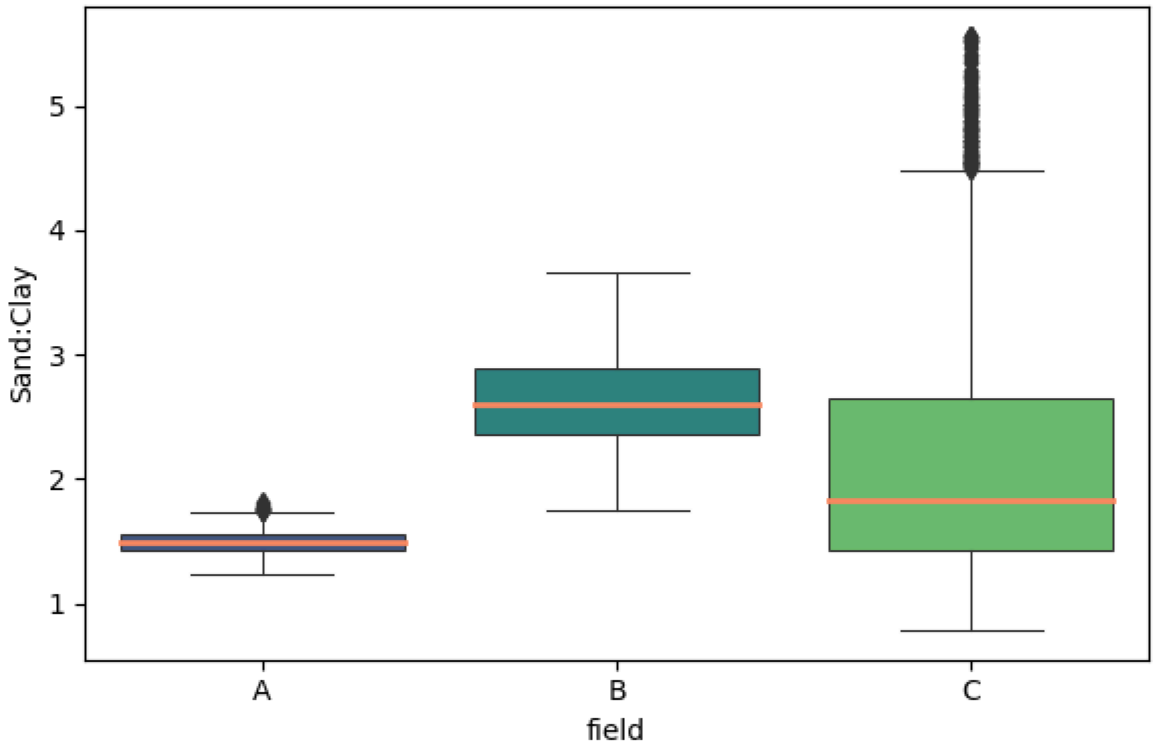

Figure 12

Soil texture variability of fields A, B and C, indicated by the sand:clay ratio and used in both the UF and SUF methods to construct zonal application maps. The orange line indicates the median.

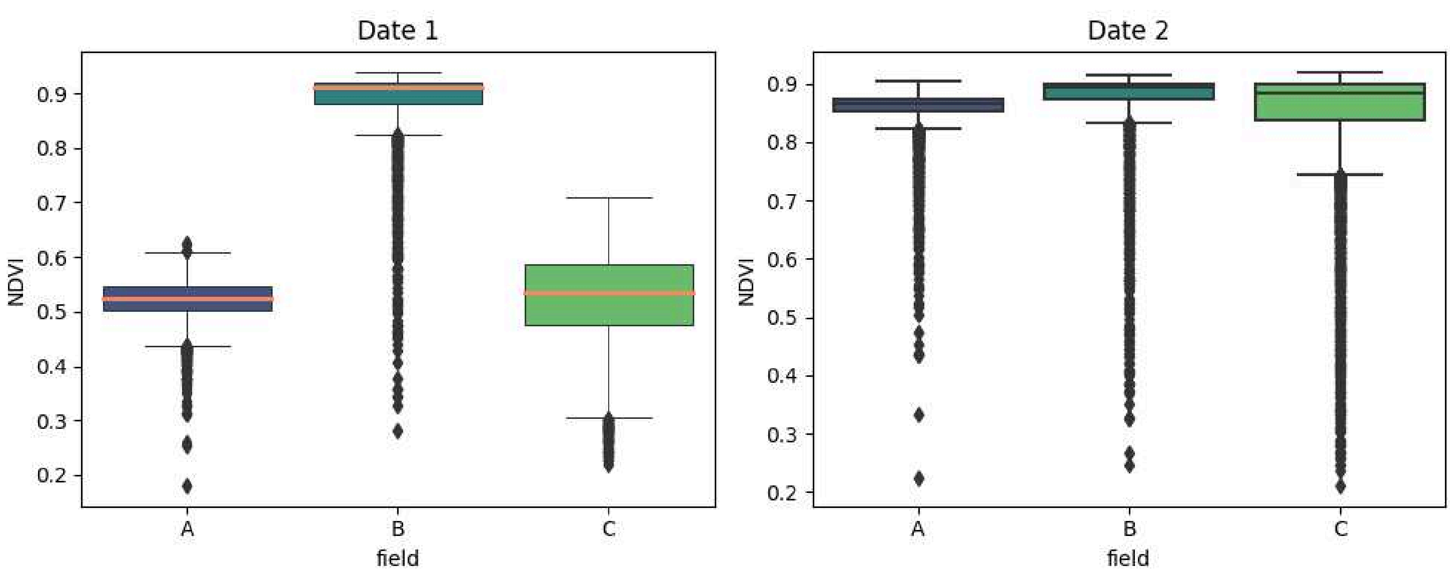

Figure 13

NDVI variability of fields A, B and C and respective dates used in the UF and SUF to construct the zonal application maps. The orange line indicates the median.

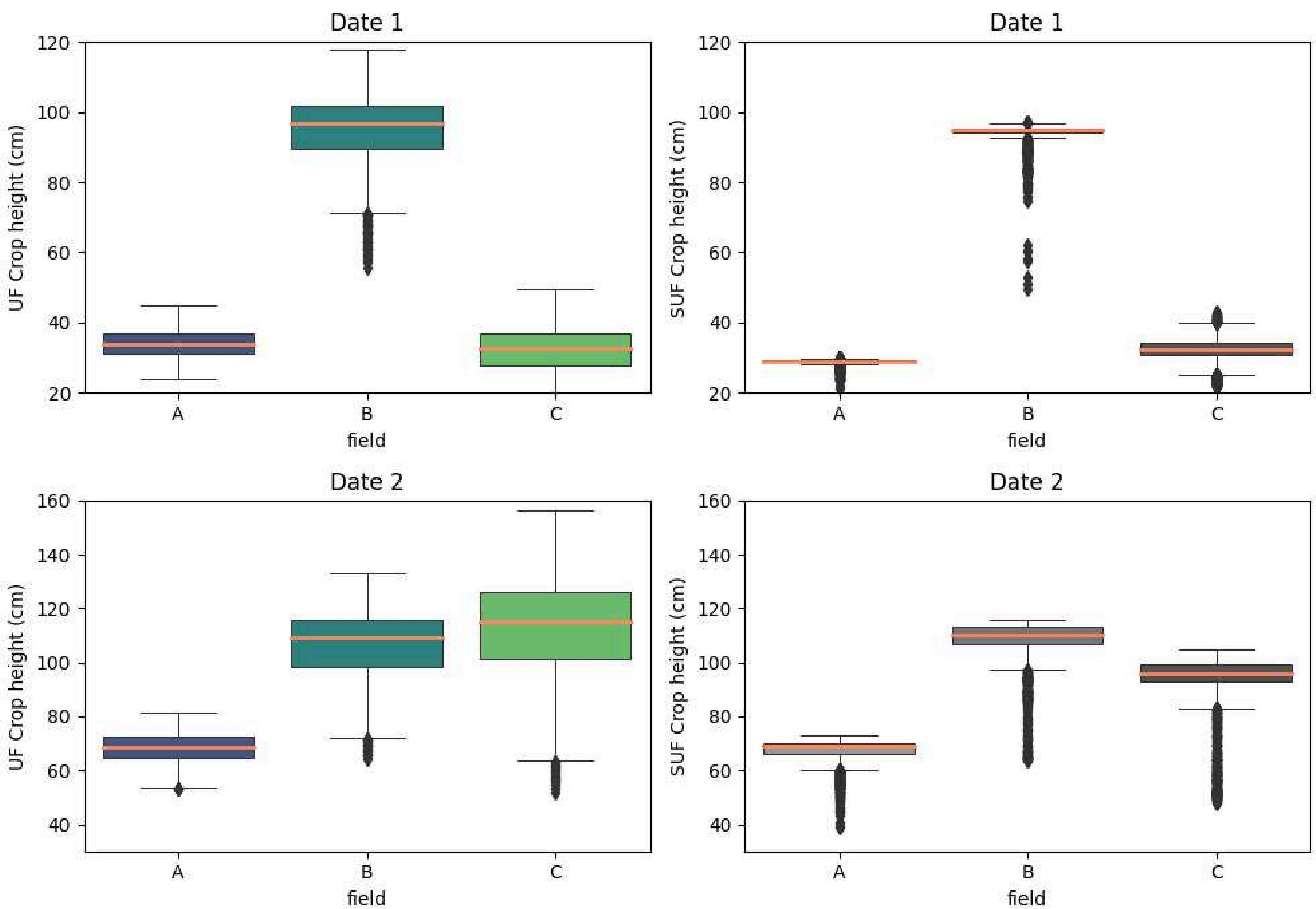

Figure 14

Crop height variability of fields A, B and C and respective dates used in the UF (plots in the left corner) and SUF (plots in the right corner) zones to construct the zonal application maps. The orange line indicates the median.

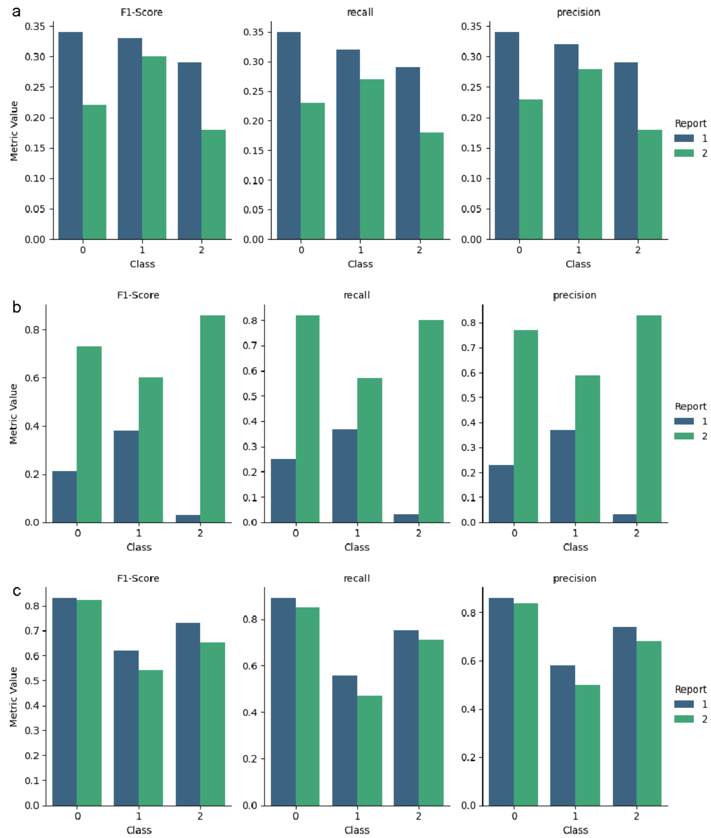

Figure 15

Metric comparison of classification reports in fields A (a), B (b) and C (c). For each field, Reports 1 and 2 refer to the first and second dates of the maps, respectively, as shown in Tables 4 and 7. Classes 0, 1 and 2 refer to the zones of application maps considering the lowest to highest crop heights, respectively.

Figure 16

Mismatch between the classification of the zones of application between UF and SUF in fields A (a), B (b) and C (c). For each field, the first and second dates refer to the first and second dates of the maps, respectively, as shown in Tables 4 and 7. The red color corresponds to mismatches between any of the classes: 0, 1 and 2.