South African Journal of Geomatics, Vol. 14. No. 2, July 2025

Mapping and Monitoring the Spatial Footprint of Agriculture in the Western Cape Province of South Africa – a Mixed Method Approach

Michael Wallace[1], Colette de Villiers[2], Annalie Fourie[3], and Cilence Munghemezulu[4,5]

[1] Western Cape Department of Agriculture, Elsenburg, South Africa
Mike.Wallace@westerncape.gov.za
[2] GIS and Data Team, Geo Pollution Technologies - Gauteng (Pty) Ltd, Pretoria, South Africa.
[3] SiQ, currently at GISCOE, Pretoria, South Africa,
[4] Department of Geography, Environmental Management and Energy Studies, University of Johannesburg, Johannesburg, South Africa,
[5] Geoinformation Science Division, Agricultural Research Council, Natural Resources and Engineering, Pretoria, South Africa

DOI: https://dx.doi.org/10.4314/sajg.v14i2.4

Abstract

This paper discusses the methodologies employed in conducting spatial agricultural censuses in the Western Cape Province of South Africa, a region characterized by its diverse agricultural landscape and the significant economic contributions of its agricultural sector. The integration of comprehensive spatial datasets into planning frameworks facilitates informed decision-making for stakeholders, researchers and regional planners, thereby contributing to the resilience of the agricultural sector amidst challenges such as climate variability, market fluctuations and competition for land resources. The Western Cape Department of Agriculture (WCDoA) thus embarked on a process to capture detailed information on agricultural activity in the province, starting in 2013 and repeated during 2017/18 and 2022/23. Production data were captured at field, orchard or vineyard scale, and in conjunction with all related data on infrastructure and agri-processing facilities. A process of a priori photographic interpretation and field mapping, followed by data gleaned from expert airborne and ground observations, was implemented, supported by remote sensing data. The 2022/23 iteration of this survey exploited recent advances in remote sensing machine learning techniques to identify annual field crops, supported by an airborne (human-observed) area frame sampling process to train and verify the algorithms applied. The resultant data, issuing from time-series census mapping, reveals spatial patterns of agricultural production, indicates the trajectories of the shifts in regional production and highlights opportunities for intensification and diversification, whilst addressing regional inequalities, to ultimately focus on and guide strategic planning. This research demonstrates successful outcomes in implementing a mixed-method approach to spatial agricultural censuses, thereby enhancing the understanding of agricultural dynamics and informing strategic responses to evolving agricultural landscapes.

Keywords: *remote sensing; sentinel-2; machine learning; winter wheat, crop classification; agricultural census; land use*

1. Introduction

1.1. The need for a spatial agricultural census in the Western Cape

Intrinsically, agriculture is a spatial production process and incorporating the spatial locations and distribution of agricultural production, has substantial logistical, market and policy implications (Senay *et al*., 2023). A key component of the effectiveness of policies and interventions aimed at improving rural well-being, agricultural development, and natural resource sustainability is our ability to adequately account for the spatial heterogeneity or disparateness, of socio-economic production and environmental conditions. Whether we are able to reliably assess and measure the spatial distribution and covariance of both socio-economic and environmental factors will determine whether we are able to make the formulation and targeting of appropriate policy and investment actions more cost-effective (You *et al*., 2014). As the agricultural sector develops, spatial census data serve as an essential resource in regional planning agencies and help to facilitate compliance with spatial planning regulations and mandated Spatial Development Frameworks (Dewar and Kiepiel, 2012). This is particularly significant in the dynamic context of agricultural commodity production, which is responsive to extrinsic factors such as changing markets, climatic shifts, water availability and competition for land resources, as is the case in a developing nation such as South Africa. Furthermore, there is value in understanding spatial patterns of agricultural production in that they might reveal untapped opportunities in, say, intensification and diversification, regional marketing, processing and trade, or might even uncover significant levels of regional inequality and as such, be helpful in shaping spatially-explicit strategic responses to such opportunities and challenges.

The Western Cape Department of Agriculture plays a pivotal role in addressing agricultural land-use issues through policy implementation and development, area-wide planning, land use management, extension services and research. The department assists in implementing sustainable practices, supports farmers with technical assistance, and ensures compliance with regulations. It also conducts research on crop and livestock production, climate adaptation, and is a source of information on agricultural land use issues. Overall, the department attempts to foster a balance between economic growth, environmental preservation, and social equity in the province's agricultural sector. Prior to 2013, there were no detailed data on land use in the province other than the broad categories resulting from the various Landsat satellite-derived projects over the years.

In order to better manage its diverse agricultural resources, the Western Cape Department of Agriculture (WCDoA) thus commenced with its first detailed airborne agricultural commodity and infrastructure census in 2013 (WCDoA, 2014). This was followed by a second iteration in 2017/18 (WCDoA, 2018) and a third during 2022/23 (WCDoA, 2024a). The third version followed a slightly different approach, making extensive use of remote sensing techniques to classify winter annual crops. Owing to the novel and technical nature of this approach, the methodology is discussed in detail in Section 2.4. These up-to-date and detailed land-use data subsequently provided a quantum shift in the level of planning data available not only to the government but also to regional planning partners and a disparate range of agricultural and conservation stakeholders (Wallace, 2021).

1.2. Study Area

As shown in Figure 1, the study area is the Western Cape Province, South Africa. This region is known for producing fruit and wine and various winter annual crops (grains and oilseeds). It extends from latitude 30.430º S to latitude 34.834º S, and from longitude 17.757º E to longitude 24.222º E. The province stands out as a prominent agricultural region for South Africa. It is highly diverse in terms of its topography, soil types and climate (see annual rainfall figures in Figure 2) and this variability dictates the need for either subtle or distinct differences in farming systems and practices in the different sub-regions (Wallace, 2018). Agriculture in the region is dependent on the rainy season, which is predominantly in the winter months, from June to August, which makes the Western Cape ideal for the planting of winter small grain and oilseed crops. The region has a wide range of landscapes, from coastal areas to highly mountainous terrain, with deeply dissected valleys, and the arid plains of the Karoo.

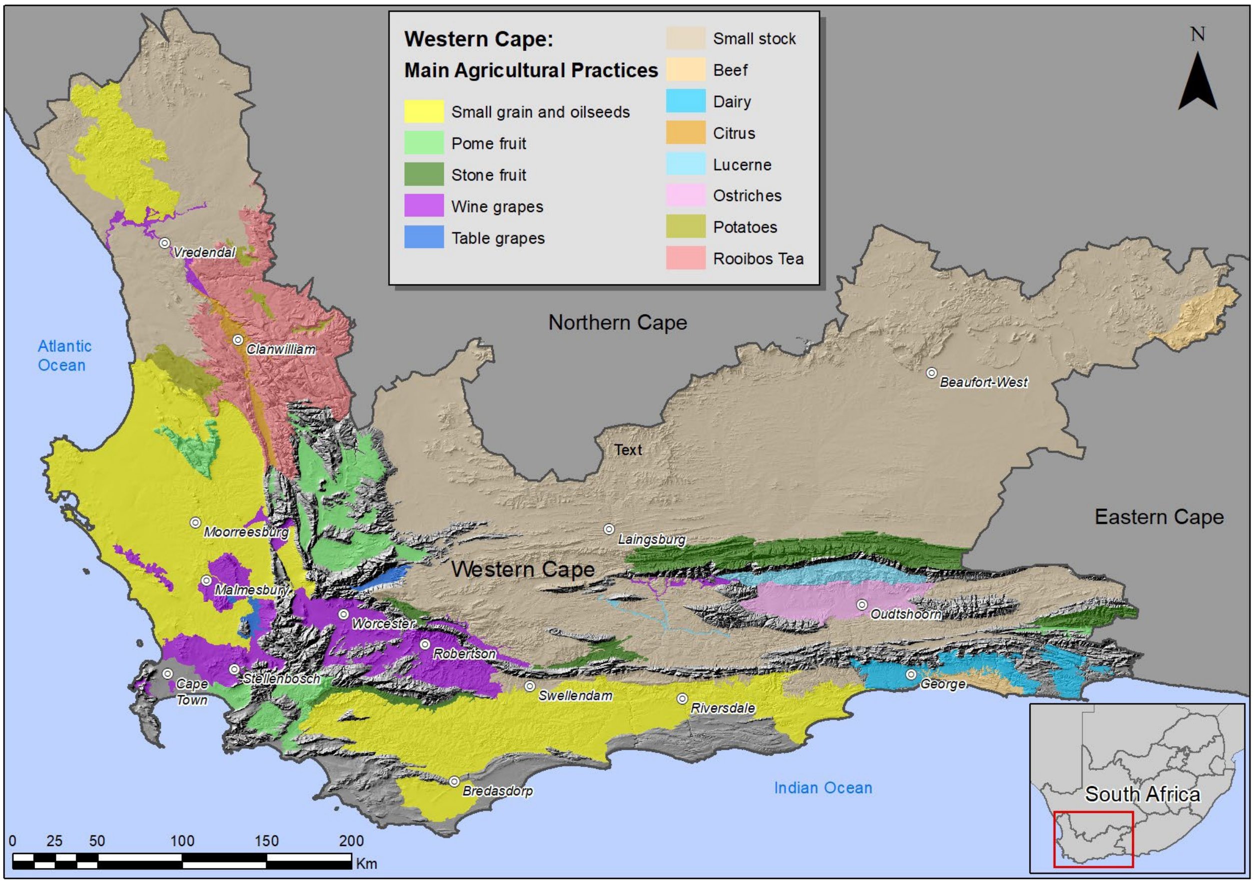

Figure 1. Western Cape province of South Africa showing the main classes of agricultural production.

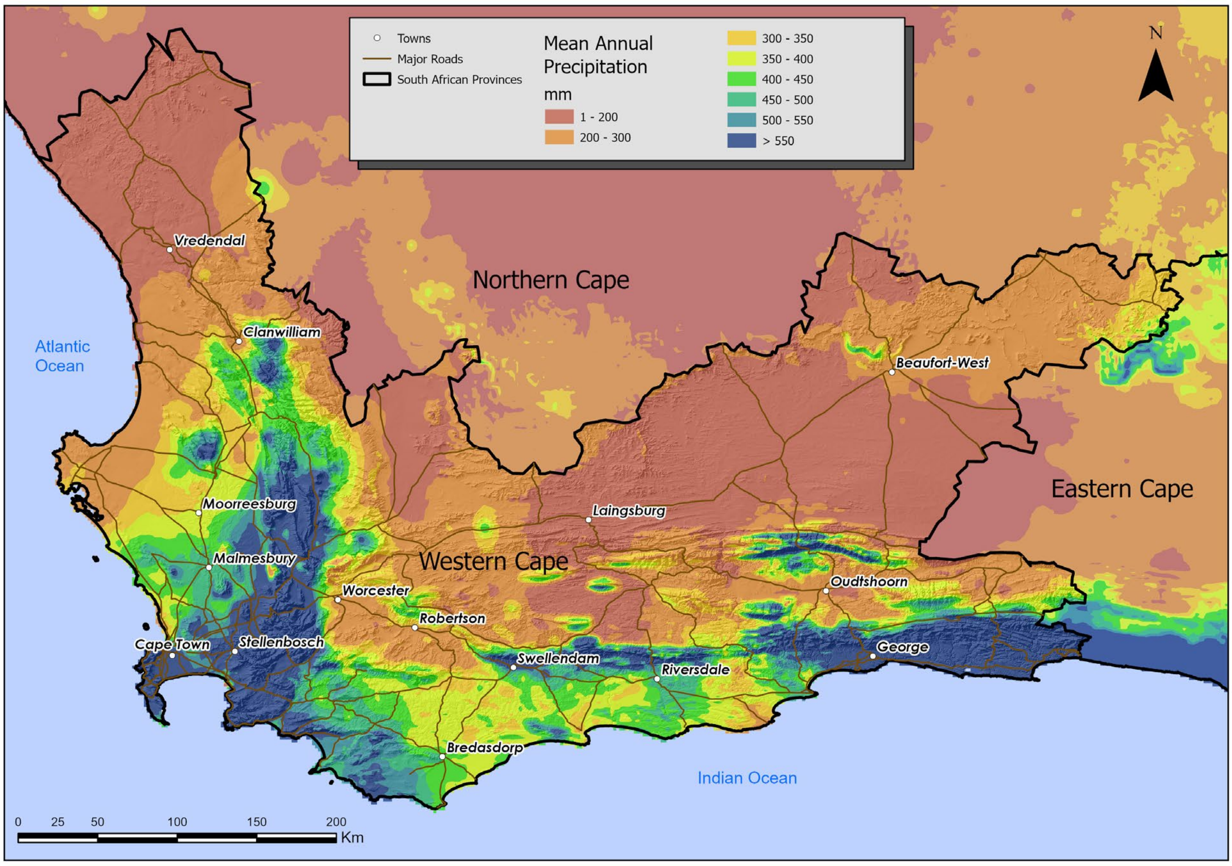

Figure 2. Western Cape province showing the range of long-term average annual precipitation. Further information on the climate and physiography of the region is available on the CapeFarmMapper Web portal (https://gis.elsenburg.com/apps/cfm/).

2. Methodology for the Agricultural Census

The methodology used for the Agricultural Census was guided by the specified final deliverables, as well as the prescribed technologies to be used, and was focused on working towards achieving the main project goals of the WCDoA as follows:

- The mapping of crop-field boundaries within the Western Cape
- Associating each field with the crop planted on the field (for the winter and summer seasons, as well as for perennial crops.
- The mapping of agriculturally-related infrastructure in the Western Cape.

The methodology can be summarised as consisting of the following processes:

- The development of an up-to-date and accurate field boundary dataset.
- An aerial survey of sample points to determine the summer annual crops planted per field.
- An aerial survey of sample points to obtain ground truthing data to be used as part of the remote sensing process for winter annual crops.

- Remote sensing to assign crop information per field for winter annual crops.
- An office exercise to assign crop information per field for horti- and viticulture, for flower, vegetable and herb production, and for various other uncommon crops.
- A vehicle-based survey to gather crop information that could not be obtained via the other processes.
- A telephonic survey to gather detailed information on the vegetable crops planted.
- An office exercise to map agriculturally-related infrastructures, such as dairies and feedlots.
- The collation of data obtained through the different processes and the implementation of quality control measures.

2.1. Field Boundary Dataset

The development of a high-quality field boundary dataset is of the utmost importance for the agricultural census. The field boundary layer is the base dataset on which census deliverables are built and provides the foundation for the accurate calculation of the hectares planted for different crops. It was therefore deemed important for the boundary of each field to be as accurate as possible.

The first step in the process was therefore to update all field boundaries in the Western Cape province. For the updating process, the latest aerial photography obtained from National Geospatial Information (NGI) was used.

The updating of field boundaries was carried out on different scales, depending on the type of area. For the horti- and viticultural areas, the scale was specified as 1:2 500, whilst the updating scale for the other areas ranged from 1:5 000 to 1:7 500.

During the updating process, field boundaries were assigned to different categories, such as, where applicable, horticulture, viticulture, pivot irrigation, small-scale farming, etc. These categories were used in various processes during the project (e.g., in the selection of sample points for the winter aerial survey and in determining the fields to be surveyed during the summer aerial survey).

2.2. Aerial Survey: Summer Annual Crops

To determine the hectares planted for summer annual crops, an aerial survey of potentially irrigated fields was conducted. The focus in this case was on irrigated fields because the Western Cape is a winter rainfall region with relatively small areas planted with summer annuals.

In preparation for the aerial survey, it was necessary to determine fields that would potentially be under irrigation but also more likely to be planted with a summer annual crop. A combination of remote sensing and field categorization techniques was used to achieve this goal:

- Firstly, *all field boundaries* that had already been categorized as horti- or viticultural during the digitizing process were excluded from the sample population as horti- and viticultural crops/fruits are perennial, are usually grown under irrigation and are generally surveyed during the comprehensive winter operations.

- For the remainder of the fields, it was decided to use the Normalized Difference Vegetation Index (NDVI) metric, derived from Sentinel-2 imagery, to identify potentially irrigated crops.

The Sentinel-2 imagery from which the NDVI values were derived was sourced for two periods. Firstly, for the start of the summer production season, and secondly, for the period at the peak stage of growth for summer crops. An average NDVI value was calculated per field for both periods.

If the average NDVI value of a field at the peak growth stage indicated dense green vegetation, the field was deemed to be planted with a crop under irrigation. From these potentially irrigated fields, the average NDVI values at the start of the summer production season were compared. If the latter's NDVI values indicated bare soil, the likelihood of the irrigated crop being a summer annual crop was deemed to be high. It was this subset of fields, with a sample point created for each field, that was included in the summer aerial survey.

An aerial survey of the sample points was conducted to determine the crop, represented by each sample point, planted on the field. The aerial observation team consisted of a pilot and a skilled observer in a helicopter. The observer was responsible for capturing the crop that had been planted on the field where the sample point was located. He was also required to indicate whether the crop was being produced under rain-fed or irrigated conditions. Data capturing was performed with a customized ArcPad interface, with the objective being for the interface to facilitate the capturing of accurate and efficient data results. The ArcPad application also included several additional datasets to assist the observer in correctly identifying the relevant fields (Fourie, 2015).

2.3. Aerial Survey: Ground Truthing Data for Winter Annual Crops

The main objective of the winter aerial survey was to obtain ground truthing data for the remote sensing process to determine the winter annual crops planted. For this purpose, statistically selected sample points, as well as additional points, were captured during the survey to increase the number of data points available for the remote sensing process.

As a preparatory measure for the aerial survey, the sample points for the survey had to be selected. The first step in the process was to design the point frame. A 45m x 45m point grid was generated over the Western Cape provincial area, with the grid points outside of the field boundaries being excluded. To increase sampling efficiency, the field boundaries were stratified according to the probability of finding a winter annual crop of interest. The stratification was based on two very important factors:

i. The categorization of the field boundaries. For instance, categories that indicate crop cultivation other than possible winter annual crops, were considered as part of the excluded strata (e.g., horticultural, viticultural and nursery plants).

ii. The density of field boundaries in an area. Field boundaries within a high-density area were classified as "high cultivation", with other density areas similarly classified as "medium cultivation" and "low cultivation".

More sample points were used in strata where there was a higher likelihood of finding winter annual crops of interest. This facilitated access to the most useful data within the budget constraints.

Sample point selection was carried out per stratum and within a database environment. The 45m x 45m grid points were imported into a database − in separate tables. It is important that sample points cover the whole of the area of interest and are evenly spread. To ensure even distributions in each case, the separate grid point tables were sorted systematically from west to east and from north to south by using the coordinates of each grid point. Within each sorted table, a random starting point was chosen. Sample points were then selected at regular intervals according to the number of points needed for a specific stratum.

For the aerial survey, the aerial observation team consisted of a pilot and two observers in a helicopter. The service of two observers was necessitated by the need to increase the number of points captured in the field over and above the sample points. The capturing was again carried out in a customized ArcPad application - the same application used for the summer aerial survey. The only difference lay in the contents of the custom interface, which would include buttons and "drop-down" lists for the winter crops as opposed to the summer crops − as was the case with the summer aerial survey.

2.4. Remote Sensing of Winter Annual Crops

During the first and second iterations of the survey, all winter crops were identified by airborne and, in some cases, ground observations, as described above. However, owing to financial constraints, the third iteration required that the number of flights of the helicopter would be reduced to constrain costs. This also provided the WCDoA with the opportunity to explore novel remote sensing techniques, particularly given the availability of free Sentinel-2 data and the advances in machine learning (ML) and artificial intelligence (WCDoA, 2024a). Prior to the census, the department had engaged with the international, not-for-profit Radiant Earth foundation (RE), an academic, community-led initiative promoting ML and AI technologies in remote sensing. An international competition was held to determine the best performing ML/AI model based on a sample of the WCDoA's 2017 crop census data. These were used as training data for the contestants (https://github.com/radiantearth/spot-the-crop-challenge). The winning models are freely available on the site and were used in the subsequent ensemble of models tested (Section 2.44).

2.4.1. Introduction to the Remote Sensing Component of the 2022/23 Census

Modern crop classification techniques can provide highly accurate agricultural monitoring results, with local governments and decision makers receiving important spatial information for improved resource allocation and food security (Bouguettaya *et al*., 2022; Karthikeyan *et al*., 2020). Large-scale crop classification is especially important in heterogeneous landscapes where to distinguish between various crops is challenging (Saini and Ghosh, 2021). The monitoring of these crops using traditional field surveys can be costly and labour intensive. Furthermore, the size of a region with crops can be difficult to cover in its full extent (Benami *et al*., 2021; Wu *et al.*, 2023).

To address these issues, this part of the study integrated remote sensing and ML techniques for more efficient crop classification. Currently, remote sensing is at the forefront of modern technologies for the monitoring and mapping of crop growth (Omia *et al*., 2023). In recent years, the satellite platform, Sentinel-2, has shown great advances in crop classifications as it provides high resolution multi-spectral imagery for modelling (Gumma *et al.*, 2022). For example, Maponya *et al.* (2020) performed a pre-harvest classification of wheat, canola, pasture, lucerne and fallow on a small scale using Sentinel-2 and machine learning with high overall accuracy (>80 %) results. This study shows that high resolution satellite and ML models can enhance the identification of crops and their classification with temporal variability.

This subsection (Section 2.4) aims to demonstrate the enhancement in satellite technologies and machine learning techniques in the mapping of winter crops on a large scale in the Western Cape Province, South Africa. The three main objectives of the winter crop mapping were to:

i. Identify different types of winter cropland (i.e. barley, canola, fallow, lucerne/medics, wheat, and oats).
ii. Evaluate the temporal accuracy of using Sentinel-2 optical imagery for mapping different winter crops.
iii. Establish the best performing machine learning crop classification method for mapping winter crops.

2.4.2. Remote Sensing Methodology

The methodology flow chart, shown in Figure 3, is the approach used in this study for mapping winter crops in the Western Cape. The methodology consists of the following: data collection and the preprocessing of satellite imagery; spectral image enhancement by creating monthly means and the computation of vegetation indices; model development and hyperparameter tuning using the GridSearch technique; crop type prediction, enabled by applying the machine learning models and training data; the testing or validation of data for accuracy assessments using testing data; model prediction; and ultimately, the creation of a final winter crop map.

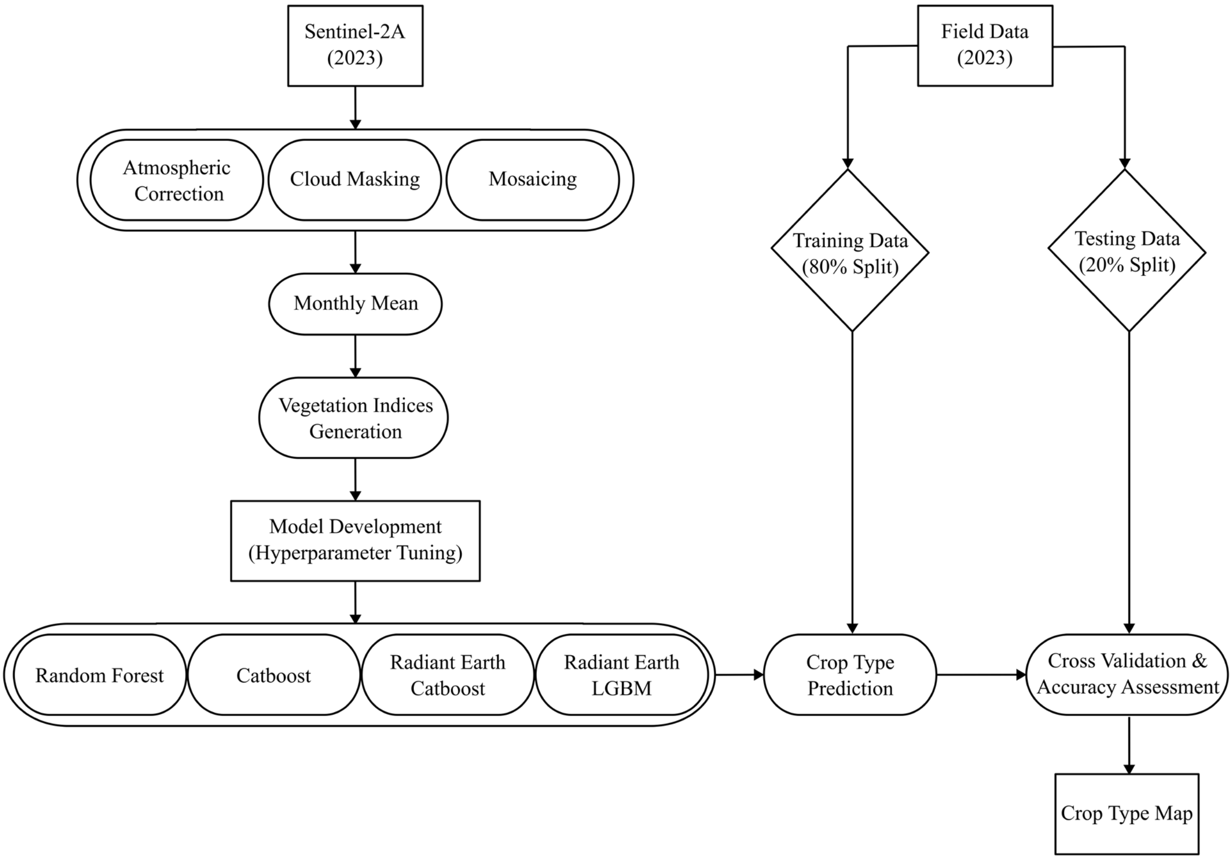

Figure 3. Overview of the methodology followed in this study.

2.4.3. Remote Sensing Data Collection and Input Features

Agricultural field data boundaries were obtained from the 2023 winter survey for the Western Cape surveyed by the service provider, SiQ (Pty) Ltd. These were used to extract sample points. From the available point data, a maximum of 2500 sampling points per class were randomly selected. This resulted in some classes having fewer sampling points for training and validation. Crop-type classes were chosen based on the field classification. Some classes were grouped together; Fallow, Weeds and Stubble were grouped into one class, as Fallow. The crop types and number of samples per class for the Western Cape Province are depicted in Table 2.

The Sentinel-2 Level-2A data were acquired from the European Space Agency (ESA) Copernicus mission. These datasets are multi-spectral satellite products with high spatial and spectral resolutions. The instruments have 13 spectral bands, including red, green, blue, near-infrared, red-edge, and shortwave infrared bands. Only seven of the 13 spectral bands available for Sentinel-2 were selected for this study. All the 10-metre resolution bands were resampled to 20-metre resolution levels using the nearest neighbour method.

The heterogeneous landscape of the Western Cape makes distinguishing features on the surface difficult. Also, the spectral similarity of crop types is often an issue. Therefore, to extract other feature spaces from the Sentinel-2 spectral bands is necessary. Table 1 lists various vegetation indices derived from the multispectral datasets.

Table 1. Description of vegetation indices derived from Sentinel-2 satellite spectral bands.

Vegetation Indices	Abbreviation	Formula	References
Anthocyanin Content Index	ACI	$NIR \times (R + G)$	Modified from Steele *et al.* (2009)
Atmospheric Resistant Vegetation Index	ARVI	$\frac{NIR - (2 \times R) + B}{NIR + (2 \times R) + B}$	Kaufman and Tanre (1992)
Ashburn vegetation Index	AVI	$\left(NIR * (1 - R) * (NIR - R)\right)$	Ashburn (1979)
Red-edge Chlorophyll Index	CLRE	$\frac{NIR}{RedEdge1 - 1}$	Gitelson *et al.* (2005)
Enhanced Vegetation Index	EVI	$2.5 \times \frac{NIR - R}{(NIR + 6R - 7.5B + 1)}$	Huete *et al.* (1999)
Excess Green Index	EXG	$2 \times \frac{G - R - B}{R + G + B}$	Woebbecke *et al.* (1995)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR - G}{NIR + G}$	Gitelson and Merzlyak (1998)
Modified Soil-adjusted Vegetation Index	MSI	$\frac{NIR - SWIR}{(NIR + SWIR + 0.5) \times (1.0 + 0.5)}$	Modified from Huete (1988)
Modified Simple Ratio	MSR	$\frac{(NIR/R) - 1}{(NIR/R)^2 + 1}$	Chen (1996)
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - R}{NIR + R}$	Rouse *et al.* (1974); Tucker (1979)
Normalized Difference Vegetation Index Red-edge	NDVIre	$\frac{NIR - RedEdge1}{NIR + RedEdge1}$	Barnes *et al.* (2000)
Normalized Difference Water Index	NDWI	$\frac{G - NIR}{G + NIR}$	McFeeters (1996)
Normalized Difference Yellowness Index	NDYI	$\frac{G - B}{G + B}$	Sulik and Long (2016)
Red-edge Normalized Difference Vegetation Index	RENDVI	$\frac{RedEdge2 - RedEdge1}{RedEdge2 + RedEdge1}$	Sims and Gamon (2002)
Red-edge Re-normalized Difference Vegetation Index	RERDVI	$\frac{NIR - RedEdge1}{\sqrt{NIR + RedEdge1}}$	Cao *et al.* (2013)
Red-edge Ratio Vegetation Index	RERVI	$\frac{NIR}{RedEdge1}$	Cao *et al.* (2013)
Soil-adjusted Vegetation Index	SAVI	$\frac{(1 + 0.5) \times (NIR - R)}{NIR + R + 0.5}$	Huete (1988)
Triangular Vegetation Index	TVI	$\frac{[120(RedEdge2 - G) - 200 * (R - G)]}{2}$	Broge and Leblanc (2001)

2.4.4. Classification Methods and Model Evaluation

The study identified four experiments (Table 2) to determine the best approach for winter crop classification. The aim for phase 2 of the project was to produce a working methodology for accurately mapping the winter crop types. Experiments 1 and 2 were developed using the April to November 2023 data, and these experiments used only the vegetation indices described in Table 3 as input features. Experiment 1 used the Random Forest classifier developed by Breiman (2001) and for this study, the scikit-learn Python library was used for processing (Pedregosa *et al*., 2011). Experiment 2 used the Catboost classifier, which is known as a gradient-boosting algorithm. Experiments 3 and 4 used data collected from April to November 2023 and followed the methodology described on the website (https://github.com/DariusTheGeek/Radiant-Earth-Spot-the-Crop-Challenge). For this study, the Pytorch and ensemble learning were omitted, and experiments 3 and 4 used machine learning techniques (Catboost and LGBM algorithms) from the Radiant Earth Spot-the-Crop challenge, that aligned well for the purpose of winter wheat classification.

Table 2. Experimental Design for four experiments conducted in this study.

Experiment	Period	Experiment Description
1	April 2023 to November 2023	RF + VIs
2	April 2023 to November 2023	Catboost + VIs
3	April 2023 to November 2023	Radiant Earth Catboost
4	April 2023 to November 2023	Radiant Earth LGBM

Classification evaluation focused on assessing model reliability and performance. First, an accuracy assessment was performed to calculate overall accuracy and user's and producer's accuracy. Second, model performance was evaluated using cross-validation to evaluate the performance on different subsets of the training and testing data.[1] Furthermore, McNemar's test[2] was applied to compare the performance of the models.

2.4.5. Results - Mapping Winter Cropping System by means of Machine Learning

For this study area, images were obtained from April to November, and the vegetation indices discussed in Table 1 were calculated. Since the winter crops flower between July and September, these were expected to be the months where winter crops could be easily identified. Four experiments were carried out to identify the best performance of the machine learning algorithms to map winter crops in the Western Cape Province. The overall accuracy of these experiments is shown in Table 3. The highest overall accuracy was observed for experiment 2 (80.31%) and experiment 3 (81.09%); these two experiments also produced the highest kappa statistics of 0.76 and 0.77, respectively. Experiment 1 produced the lowest overall accuracy (75.81%) and kappa statistic (0.71). The results

---

[1] This is essential to determine whether the model performs favourably without the overfitting of data.

[2] This is a statistical test used to compare the statistical significance among different models.

shown here indicate that the Catboost algorithm produces better accuracy models than the other machine learning algorithms.

Table 3. The overall accuracy and kappa coefficients for the four experiments using July-to-September imagery.

Experiment	Experiment Description	Image Dates	Overall Accuracy (%)	Kappa
1	RF	2023/04-2023/11	75.81	0.71
2	Catboost	2023/04-2023/11	80.32	0.76
3	Radiant Earth Catboost	2023/04-2023/11	81.09	0.77
4	Radiant Earth LGBM	2023/04-2023/11	79.18	0.75

Tables 4 to 7 present the confusion matrices for the four experiments (1 − 4), which provide valuable insights into the performance of each experiment for the respective crop types. The confusion matrix in each table summarizes the predicted and actual classifications, with the diagonal values indicating the correctly classified pixels. These matrices highlight the performance of each model by identifying errors within each of the models. The accuracy metrics provided in Table 3 were calculated from the results of these confusion matrices.

Table 4. The confusion matrix for crop classification in Experiment 1.

	Barley	Canola	Fallow	Lucerne/ Medics	Oats	Planted Pastures	Wheat
Barley	**143**	14	24	14	10	0	22
Canola	6	**439**	8	14	15	0	12
Fallow	1	4	**651**	56	11	13	4
Lucerne/ Medics	2	9	95	**466**	26	14	8
Oats	14	15	34	44	**204**	7	72
Planted Pastures	0	7	37	54	20	**94**	5
Wheat	10	5	22	12	21	0	**388**

Table 5. The confusion matrix for crop classification in Experiment 2.

	Barley	Canola	Fallow	Lucerne/ Medics	Oats	Planted Pastures	Wheat
Barley	**205**	4	8	6	15	0	9
Canola	3	**446**	7	10	10	1	9
Fallow	1	2	**685**	61	9	11	0
Lucerne/ Medics	5	0	66	**465**	20	16	9
Oats	19	4	28	40	**219**	10	50
Planted Pastures	2	1	31	60	27	**99**	3
Wheat	9	3	13	7	30	0	**408**

Table 6. The confusion matrix for crop classification in Experiment 3.

	Barley	Canola	Fallow	Lucerne/ Medics	Oats	Planted Pastures	Wheat
Barley	**204**	4	10	7	14	1	7
Canola	5	**447**	6	10	6	3	9
Fallow	1	0	**692**	54	9	13	0
Lucerne/ Medics	3	1	63	**466**	20	18	10
Oats	18	2	27	39	**222**	12	50
Planted Pastures	2	1	27	54	24	**113**	2
Wheat	10	5	11	4	33	0	**407**

Table 7. The confusion matrix for crop classification in Experiment 4.

	Barley	Canola	Fallow	Lucerne/ Medics	Oats	Planted Pastures	Wheat
Barley	**192**	5	11	9	17	0	13
Canola	4	**442**	7	14	8	5	6
Fallow	0	0	**675**	71	7	15	1
Lucerne/ Medics	5	1	73	**458**	23	12	9
Oats	20	3	37	37	**219**	8	46
Planted Pastures	1	1	25	72	26	**97**	1
Wheat	8	4	11	5	34	0	**408**

The results from the cross-validation − obtained to further emphasize the accuracy of the models − are shown in Figure 4. Experiment 4 showed the highest cross-validation accuracy scores, with the median value above 0.825, indicating that the model accurately trained the training samples. With median scores close to 0.82, the boxplots and whiskers plots for experiments 2 and 3 show that their accuracies consistently indicate similar accuracies. These findings offer high confidence levels as to the accuracy of experiments 2, 3 and 4. Unfortunately, with scores below 0.77, experiment 1 produced less favourable cross-validation accuracy scores. The median for experiment 1 was also indicated as below 0.76. These findings indicate that, overall, the models performed favourably for experiments 2, 3 and 4, whereas the models created in experiment 1 performed less favourably.

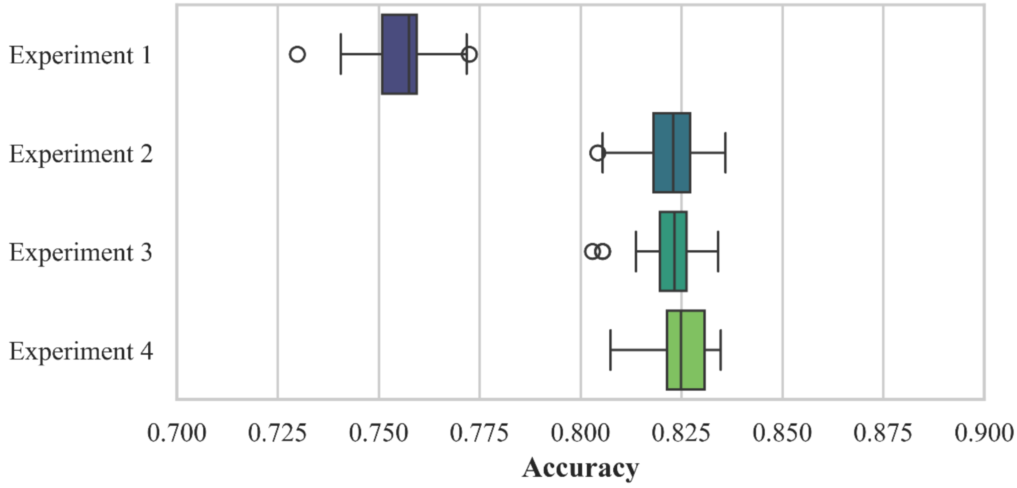

Figure 4. The cross-validation accuracy scores for the four experiments. The boxplots are the lower and upper quartiles (whiskers in black) and the black line in the boxplot is the median. The circles represent outliers in the data outputs, which are data points outside the lower and upper quartiles.

Figure 5 represents the accuracies for experiments 1, 2, 3, and 4 across six different classes (Barley, Canola, Fallow, Lucerne/Medics, Planted Pastures, Oats, and Wheat). From these plots, it is evident that the models have varying levels of performance across the respective crop classes. Among the experiments, experiment 2 (Catboost), experiment 3 (Radiant Earth Catboost) and experiment 4 (Radiant Earth LGBM) consistently achieved high precision and recall levels and F1-scores across most classes, while experiment 1 (Random Forest) produced the lowest classification metrics throughout all the classes. In terms of performance per class, Planted Pastures represents the lowest-performing class across all models – likely a consequence of the inherent heterogeneity of the class. Canola proved to be the highest performing class, with all four experiments consistently delivering the best metrics of precision, recall and F1-scores for this class.

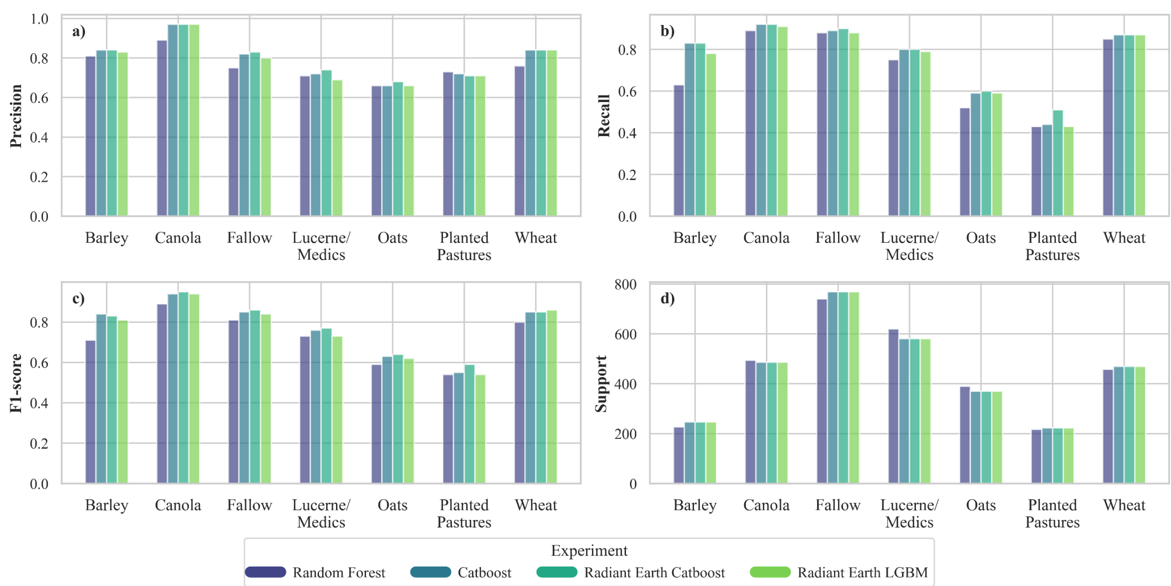

Figure 5. The bar graphs for per-class precision, recall, F1-score and support metrics for each of the experiments/classifications.

2.4.6. McNemar's Test of Statistical Significance

Table 11 presents the results of the McNemar's test for comparing the performance of the respective experiments. This is a statistical test used to calculate the significant difference between two paired models and is based on the confusion matrices determined during crop classification.

A comparison of the experiment 3 results with those of the other three models showed no statistically significant differences in the performance of the models. The p-values exceeded the significance level of 0.05 in all three comparisons with experiment 3. This indicates that there is no strong evidence to suggest that one experiment performed significantly better than the other. The Z-values also consistently reflected minimal differences in performance amongst the models. These findings suggest that, within the context of the data and evaluation metrics, the models exhibit similar performance levels, and no clear advantage is to be found in choosing one model over another.

Table 11. McNemar's Test Results for Model Comparisons

	Z-value	p-value
Experiment 1 *vs* Experiment 3	0.3162	0.7518
Experiment 2 *vs* Experiment 3	0	1
Experiment 4 *vs* Experiment 3	0.3536	0.7237

2.4.7. Assessing the Spatial Patterns of Crop Types and Field Mapping for the Winter Crop using the Machine learning Method

The crop-type map representing the Western Cape Province was created using the model from experiment 2 (Catboost). The winter crop map indicates that wheat is the dominant crop class in the western central region of the province (Figure 6A). Towards the southern to southeastern regions of the province where a large section of cropland is located, a more diverse variety of crop types was mapped (Figure 6B). The results reveal that the fields were generally mapped homogeneously within each field boundary, with few of the field boundaries embracing more than one crop type, although this was not always the case. Mixing, also known as pixel mixing, occurs throughout for different reasons (e.g., along the field edge/boundary and where features or crop types have similar spectral signatures). Environmental factors and crop-health issues could also influence field homogeneity, thereby causing inconsistencies in the field.

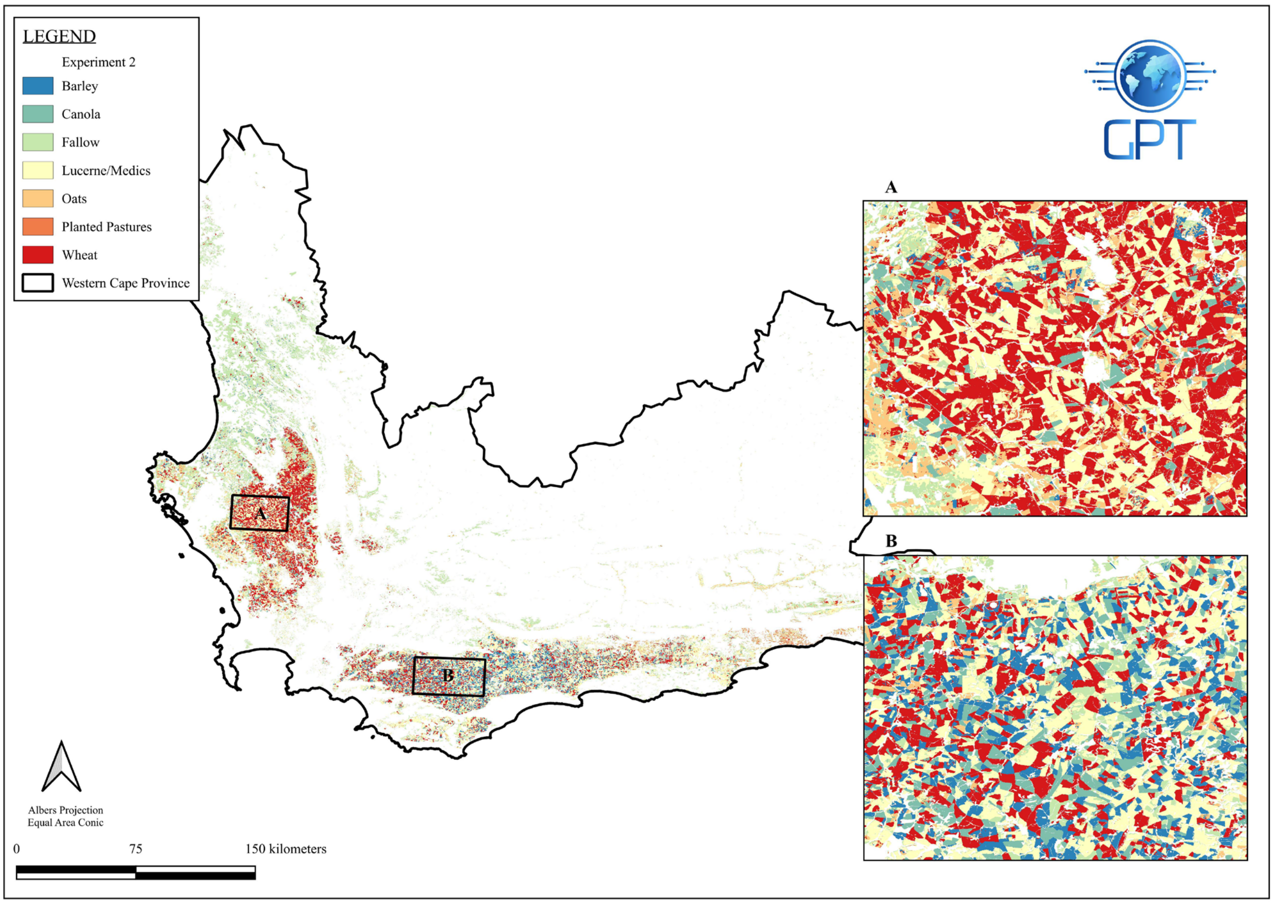

Figure 6. Winter crop type distribution from experiment 2 for the Western Cape.

2.4.8. Concluding Remarks on the Remote Sensing Approach

The findings of this study suggest that the Catboost machine learning methods produced the highest accuracy results − above 81% overall accuracy for experiment 3. The results of experiment 4 show only a two percent (2%) difference in accuracy compared to those of experiment 3, and a one percent (1%) difference from those of experiment 2, with the results from experiment 1 (RF) showing the lowest classification accuracy, with a five percent (5%) difference in overall accuracy as opposed to those of experiment 3. The statistical significance of McNemar's Test shows that the results of these experiments produced no significant statistical difference between the performance of the models. This indicates that, based on this test, no specific experiment can be identified as being significantly superior in accuracy over another. However, based on the overall accuracies of the models, the model in experiment 3 was chosen as the final classification for winter crops in 2023. It is noteworthy that the other experiments (experiments 2 and 4) also demonstrate success in accurately mapping the winter crops for 2023. The following classes were successfully identified: barley, canola, fallow, lucerne/medics, planted pastures, wheat, and oats. The greatest confusion arose in the planted pastures class. This is generally caused by the concept of pixel mixing, which occurs when the different classes are confused, primarily in cases where the spectral signatures are similar for two or more classes. This could also be due to the smaller number of available ground truth data points

2.5. Assignment of Crop Information for Other Crop Types

Crops other than winter annual crops and pastures, that is for horti- and viticulture, flowers, vegetables and herbs, were excluded from the remote sensing process. For these crops, a different approach was necessary to assign the crops to fields.

An office classification process was followed and can be summarized as follows:

- Firstly, all fields to which these crops were assigned during the previous census of 2017/2018 were extracted.
- For these fields, the crop indicated on the field during the 2017/2018 census was compared with the current crop planted, as observed on up-to-date Google Earth imagery, as well as on aerial photographs.
- Should it be found that the crop observed on the indicated imagery is the same as the one planted previously, during the 2017/2018 census, the applicable crop would again be assigned to the field.
- Horti- and viticulture where there was a possible or definite change in crop, or other discrepancies (e.g., crops that were previously incorrectly classified), were selected to be included in a vehicle-based survey. Fields that were potentially planted with new flower crops, vegetables and herbs, were also included in the indicated survey.

2.6. Vehicle-based Survey

The objective of the vehicle-based survey was to gather information that could not be obtained via other means. The information that had to be collected included horti- and viticultural crops which could not be categorized via the office classification process, as well as those under shade-netting and in tunnels where the crop planted was unknown. For the vegetable survey, it was also necessary to obtain the contact details of the vegetable farmers in cases where no or incorrect observations for these farms were available. This was one of the few instances where farmers were contacted directly.

Data capturing for the vehicle-based survey was conducted in ArcGIS. The capturing interface that was set up used a Bluetooth GPS to assist the field surveyor in identifying the fields to be captured within a moving map display. To facilitate accurate and efficient data capturing, templates were set up for all the datasets used in collecting data.

2.7. Vegetable Telephonic Survey

A telephonic survey of all vegetable farmers whose contact details were available and valid was conducted. The objective of the survey was to obtain production figures from the producers concerning the vegetables that had been planted during the year. Owing to the relatively rapid within-season crop rotations applicable to vegetable production, this was necessary since it is difficult to gain accurate figures on the number of hectares planted under vegetables when other techniques are used.

The farmer survey included questions on the number of hectares planted, the yield and production per vegetable crop, with separate sets of figures provided for vegetables planted on open fields, under shade-netting, in tunnels and according to other methods.[3] Information on crop rotation practices was also gathered.

2.8. Mapping of Agriculturally-related Infrastructures

An important part of the census project was the capturing and verification of livestock infrastructure and agri-processing facilities. The livestock infrastructure captured included information on abattoirs and livestock auction facilities, aquaculture, chicken batteries, dairies, feedlots and piggeries, as well as other livestock infrastructures (e.g., goat pens and kraals for sheep and cattle). With regard to agri-processing, all types of cellars were captured, as well as other types of facilities (e.g., packhouses, cool chain, milling and tea processing facilities).

The process followed, focused on the verification of infrastructures captured during the previous census projects, as well as the capturing of new infrastructures. The same aerial photographs were used as for the updating of the field boundaries. For the accurate identification of the infrastructures to be mapped, it was necessary to zoom in on aerial photographs at a scale of approximately 1:1000.

Various sources were used for the verification of the infrastructures. These included the internet, Google Earth, Google Maps, as well as Google Street View. The verification process entailed verifying the location, as well as the existence of an agri-processing facility, where applicable. If it could be obtained, it was also necessary to capture attribute data for agri-processing facilities, such as the type of facility, the name of the facility, as well as the contact details for the facility.

2.9. Data Collation and Quality Control

The information gathered during the various surveys and processes was processed and comprehensive quality assurance was performed on all datasets. The specific quality control procedures followed, varied among the respective datasets and were set up to be the most appropriate for each dataset. An important part of the quality assurance that was carried out, was cross-checking among the different datasets. Specifically, the field boundaries and agriculturally-related infrastructures, as well as the aerial survey, the vehicle-based survey and the crop assignment process, contributed to this process.

The respective sources of field-based information were combined to create a comprehensive GIS dataset that also included field boundaries, with the applicable crop assigned to each field. A separate dataset was developed from the summer aerial survey, to provide information on the planted summer

[3] "Other methods" in this context refers mainly to the production of mushrooms and *waterblommetjies*.

annual crops. Information gathered during the telephonic vegetable survey was processed and linked to farming operation boundaries, thereby providing a spatial context for the annual vegetable production figures.

3. Results and Open Access to the Data

The resultant primary database developed in each of the iterations contains detailed information (for each of the winter and summer operations) regarding crop type, various commodity sub-classifications, a field to indicate whether the field/orchard/vineyard appears to be under irrigation, date of each record, imagery used to digitise the field boundary (*a priori*), mode of attribute observation, observer name, local municipal area into which the record falls and the area of each commodity record. Together with these details, is a spatial database of all agricultural infrastructures, such as abattoirs, wine cellars, grain storage facilities, shade-netting, tunnels, dairies, piggeries and various other agri-processing facilities.

During the 2022/23 iteration, game farms were also mapped with the assistance of Cape Nature and South African National Biodiversity Institute (SANBI). Also amongst the requested 2022/23 deliverables were a methodology report (WCDoA, 2024a) and a strategic analysis of the data to inform the department and provincial planning partners (WCDoA, 2024b).

Agricultural land use is one of the most requested datasets from the WCDoA's Geographic Information System (GIS) Unit (Wallace, 2013). It was thus planned at the outset to make the census data available on an online GIS viewer application. The department developed an open user portal, known as CapeFarmMapper (Basson, 2013), whereby the census datasets described above could be viewed (Figure 7) and analysed in conjunction with various other spatial layers (https://gis.elsenburg.com/apps/cfm/). The platform has proved to be extremely popular and attracts thousands of users per month – many of whom are interacting with the agricultural commodity and infrastructure datasets (Wallace, 2021).

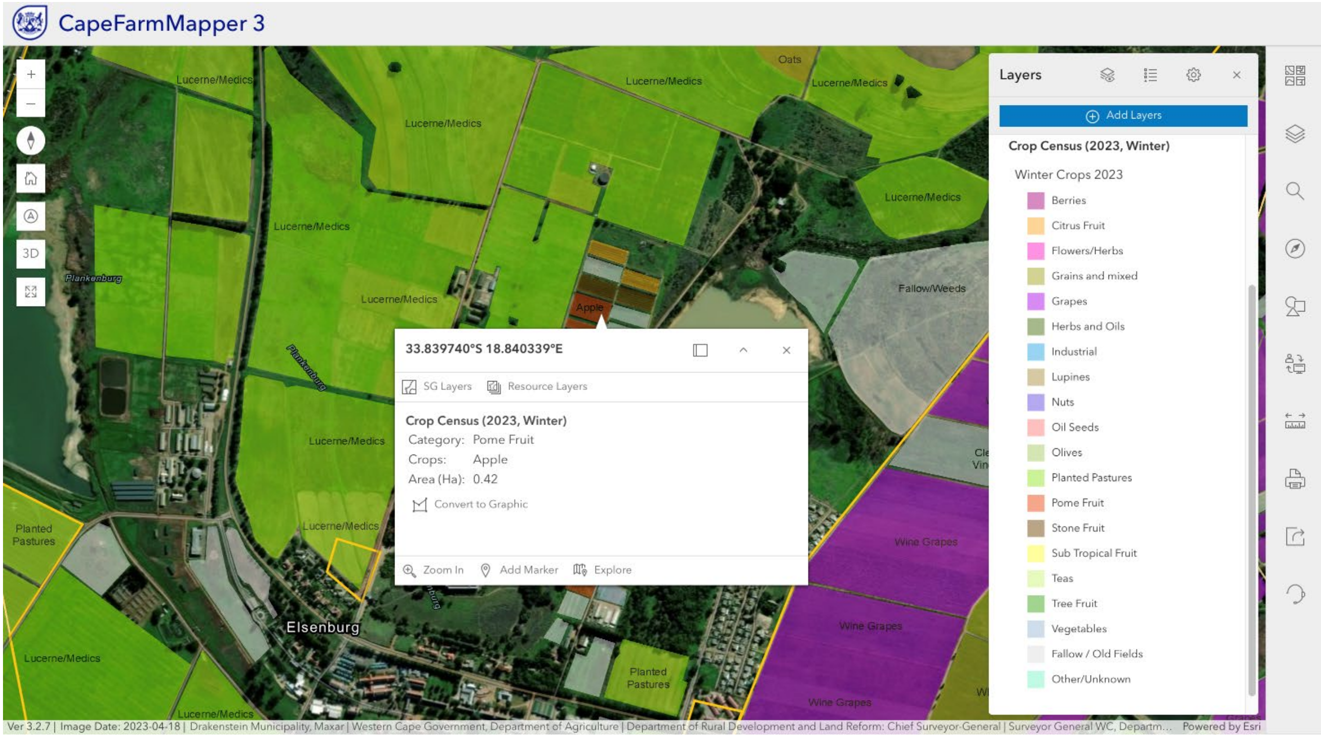

Figure 7. A screenshot showing the level of detail in the census land use data.

4. Conclusion

The enormity of the various iterations in this census project is not to be underestimated. The number of crop fields/orchards/vineyards in the datasets approached 300 000, together with the many thousands of infrastructural points also captured, with their attributes. Since the 2013 census, various methodologies have been developed and refined over time to most effectively and efficiently capture the data required by WCDoA and its stakeholders. These range from attribute annotations based on skilled airborne and ground observations to intensive Web-based research, the analysis of Google Maps and Street View imagery, and, where absolutely necessary, telephonic queries. The process was also supported by local intelligence gathering and valued inputs following on many consultations with commodity groups and organised agriculture. The third iteration, providing satisfactory results, broke new ground in applying complex machine learning algorithms to classify a time series of satellite imagery gathered through the winter growing season.

The considerable advantage of gathering such detailed census data in the spatial context is that it can provide regional statistics for any zone, or facilitate analysis at its most granular level for on-farm planning and operational interests. These data have provided not only industry stakeholders with valuable local information, but in the context of developing countries, new standards for government town and regional planners to integrate detailed and easily accessible land-use data into their strategic decision-making at all levels and across many disciplines. The repetition of the survey three times over a decade has also provided some unprecedented insights into the trajectories of change as producers innovatively adjust to market forces, climate variability, water availability and the ever-increasing competition for land resources.

5. References

Ashburn, P., 1979. The vegetative index number and crop identification. *NASA. Johnson Space Center: Proceedings of Technical Sessions*, Vol. 1 and 2.

Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data, *Proceedings of the Fifth International Conference on Precision Agriculture*, Bloomington, MN, USA, p. 6.

Basson, F., 2013. CapeFarmMapper: Spatial data at your fingertips. *AgriProbe,* 18, 2.

Benami, E., Jin, Z., Carter, M.R., Ghosh, A., Hijmans, R.J., Hobbs, A., Kenduiywo, B., Lobell, D.B., 2021. Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews: Earth & Environment,* 2, 140-159.

Bouguettaya, A., Zarzour, H., Kechida, A., Taberkit, A.M., 2022. Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Computing and Applications,* 34, 9511-9536.

Breiman, L., 2001. Random Forests. *Machine Learning,* 45, 5-32.

Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment,* 76, 156-172.

Cao, Q., Miao, Y., Wang, H., Huang, S., Cheng, S., Khosla, R., Jiang, R., 2013. Non-destructive estimation of rice plant nitrogen status with crop circle multispectral active canopy sensor. *Field Crops Research,* 154, 133-144.

Chen, J.M., 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing,* 22, 229-242.

Dewar, D., Kiepiel, J., 2012. Spatial development frameworks on a broader scale: an integrative approach. *Town and Regional Planning,* 2012.

Fourie, A., 2015. GIS in Agriculture assists Informed Decision Making, PositionIT. EE Publishers, Muldersdrift, South Africa.

Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research,* 22, 689-692.

Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters,* 32.

Gumma, M.K., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, R.N., Whitbread, A.M., 2022. Crop type identification and spatial mapping using Sentinel-2 satellite data with focus on field-level information. *Geocarto International,* 37, 1833-1849.

Huete, A., Justice, C., Van Leeuwen, W., 1999. MODIS vegetation index (MOD13). *Algorithm Theoretical Basis Document,* 3, 295-309.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment,* 25, 295-309.

Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology,* 586, 124905.

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing,* 30, 261-270.

Maponya, M.G., Van Niekerk, A., Mashimbye, Z.E., 2020. Pre-harvest classification of crop types using a Sentinel-2 time series and machine learning. *Computers and Electronics in Agriculture,* 169, 105164.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing,* 17, 1425-1432.

Omia, E., Bae, H., Park, E., Kim, M.S., Baek, I., Kabenge, I., Cho, B.-K., 2023. Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sensing,* 15, 354.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research,* 12, 2825-2830.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication,* 351, 309.

Saini, R., Ghosh, S.K., 2021. Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. *Geocarto International,* 36, 2141-2159.

Senay, S.D., Greyling, J.C., Pardey, P.G., Verhoef, H., 2023. Spatializing South African agricultural censuses, 1918–2017. *Agrekon,* 62:3-4, 328-336.

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment,* 81, 337-354.

Steele, M.R., Gitelson, A.A., Rundquist, D.C., Merzlyak, M.N., 2009. Nondestructive estimation of anthocyanin content in grapevine leaves. *American Journal of Enology and Viticulture,* 60, 87-92.

Sulik, J.J., Long, D.S., 2016. Spectral considerations for modeling yield of canola. *Remote Sensing of Environment,* 184, 161-174.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment,* 8, 127-150.

Wallace, M., 2013. Spatial Intelligence. Mapping of agricultural commodites and infrastructure in the Western Cape. *AgriProbe,* 10, 2.

Wallace, M., 2018. Spatializing the modelled impacts of future climate change – a case study on wheat in the Western Cape. *South African Journal of Geomatics,* Vol. 7. Part 2, 9.

Wallace, M., 2021. CapeFarmMapper continues to grow! *AgriProbe,* 18, 48-51.

WCDoA, 2014. The Mapping of Agricultural Commodity Production and Agricultural Infrastructure in the Western Cape Province, in: SiQ (Ed.). *Western Cape Department of Agriculture*, Elsenburg, Western Cape, p. 117.

WCDoA, 2018. Mapping of Agricultural Commodities and Infrastructure in the Western Cape, in: BFAP, SiQ (Eds.). *Western Cape Department of Agriculture*, Elsenburg, p. 44.

WCDoA, 2024a. The Mapping of Agricultural Commodity Production and Infrastructure in the Western Cape Province: Part 1, in: SiQ (Ed.). *Western Cape Department of Agriculture*, Elsenburg, p. 75.

WCDoA, 2024b. The Mapping of Agricultural Commodity Production and Infrastructure in the Western Cape Province: Part 2 - Strategic Report, in: OABS (Ed.). *Western Cape Department of Agriculture*, Elsenburg, p. 78.

Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A., 1995. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE,* 38, 259-269.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A.B., Qin, X., Yan, N., Chang, S., Zhao, Y., Dong, Q., 2023. Challenges and opportunities in remote sensing-based crop monitoring: A review. *National Science Review,* 10, nwac290.

You, L., Wood, S., Wood-Sichra, U., Wu, W., 2014. Generating global crop distribution maps: From census to grid. *Agricultural Systems,* 127, 53-60.