Data Descriptor

Unmanned Aerial Vehicle (UAV) and Spectral Datasets in South Africa for Precision Agriculture

Cilence Munghemezulu *, Zinhle Mashaba-Munghemezulu, Phathutshedzo Eugene Ratshiedana, Eric Economon, George Chirima and Sipho Sibanda

Abstract: Remote sensing data play a crucial role in precision agriculture and natural resource monitoring. The use of unmanned aerial vehicles (UAVs) can provide solutions to challenges faced by farmers and natural resource managers due to its high spatial resolution and flexibility compared to satellite remote sensing. This paper presents UAV and spectral datasets collected from different provinces in South Africa, covering different crops at the farm level as well as natural resources. UAV datasets consist of five multispectral bands corrected for atmospheric effects using the PIX4D mapper software to produce surface reflectance images. The spectral datasets are filtered using a Savitzky–Golay filter, corrected for Multiplicative Scatter Correction (MSC). The first and second derivatives and the Continuous Wavelet Transform (CWT) spectra are also calculated. These datasets can provide baseline information for developing solutions for precision agriculture and natural resource challenges. For example, UAV and spectral data of different crop fields captured at spatial and temporal resolutions can contribute towards calibrating satellite images, thus improving the accuracy of the derived satellite products.

Dataset: Available on request to munghemezuluc@arc.agric.za. The ARC (Agricultural Research Council) is in the process of developing an interface that will allow users to have access to the data directly from our server in line with our policies on data sharing. Therefore, in the meantime, users can request the data.

Dataset License: CC BY-NC 4.0

Keywords: unmanned aerial vehicles; spectral data; precision agriculture; high-resolution imagery

1. Summary

Remote sensing datasets have been widely used in agriculture [1–3] and natural resource management [4] applications. Most publicly available datasets are of a coarse-to-medium spatial resolution; examples of these are Landsat-9 [5], Sentinel-1, and Sentinel-2 [6] datasets. Consequently, coarse-resolution imageries are unable to resolve fine-scale land features which reduces the reliability of satellite-derived products for natural resource interventions [7]. Apart from resolution issues, freely available satellite products are available on fixed temporal resolutions [8] which might not be ideal for immediate decision making. Higher-resolution imagery is available from commercial companies such as PlanetScope. However, these datasets can be very expensive. With the development of unmanned aerial vehicles (UAVs) in recent years, it is now relatively affordable to use UAVs for applications such as precision agriculture and natural resource management with ultra-high spatial resolution at the centimeter level. The application of UAVs in mapping natural resources offers flexibility in data acquisition within a flexible user-demand time frame provided the conditions are permissible [9].
Applications of UAVs in natural resource management are the focus of this paper. However, UAVs have proven to be useful in other fields besides agriculture. UAVs enable the collection of valuable data for environmental research, such as monitoring wildlife populations [10], assessing ecosystem health [11], mapping vegetation [12], and tracking changes in land use [13]. They play a crucial role in disaster management by assisting in search and rescue operations, assessing damage after natural disasters, and delivering emergency supplies to affected areas [14]. UAVs equipped with advanced sensors and Global Positioning System (GPS) technology can quickly and accurately map terrain, create topographic surveys, and gather geographical data for urban planning, mining, and land management [15].

The advantages of UAVs are that they provide a relatively affordable solution for data collection compared to traditional methods involving satellites, manned aircraft, or ground-based surveys. UAVs can be deployed and maneuvered easily in various environments, offering flexibility in monitoring different terrain or locations. High-resolution cameras and advanced sensors onboard UAVs allow for precise data collection, resulting in accurate mapping, monitoring, and analysis. Despite these advantages, UAVs have limitations; flight time is limited by battery constraints, necessitating careful planning and potentially multiple flights for comprehensive monitoring. Adverse weather conditions such as strong winds, rain, or fog can impede drone operations and affect data quality. Managing and processing large volumes of data collected by UAVs may require specialized software, skilled personnel, and computational resources. Most farmers have no access to UAV systems. This is due to a lack of resources and high levels of governmental regulation. Several Agricultural Research Council-Natural Resources and Engineering (ARC-NRE) projects in South Africa have introduced the technology to support smallholders and emerging farmers. Such initiatives ensure knowledge transfer and provide access to digital information to improve crop production. The ARC-NRE aims at identifying problems hindering the productivity of smallholders and emerging farmers. Different projects focusing on soil assessments, water resource investigations, nutritional status, and disease identification using geospatial technologies are constantly being carried out. The availability of high-resolution data allows crop scientists and researchers to understand crop–soil–water variabilities from a plant level to a field scale in detail as opposed to the use of publicly available satellite information. High-resolution information is necessary for model development, calibration, and validation at different scales emerging from local scales.

The aim of this paper is to describe the UAV and spectral datasets collected from various projects at the ARC-NRE. The datasets have been collected over a number of years from different funded projects across South Africa (Figure 1, Table 1). Most study locations are within farming areas, covering different crops such as maize, wheat, and barley. A standard data processing procedure was followed to pre-process and generate surface reflectance images from the UAV system by using Pix4D mapper software version 4.8.4 [16]. The software has been widely used by researchers to pre-process UAV images, for example, Su et al. [17]; Dimyati et al. [18]; Song et al. [19]; and Chaudhry et al. [20]. An Analytical Spectral Device (ASD) Fieldspec spectroradiometer instrument was also used to collect spectral reflectance data of different crops in fields. Clevers et al. [21] and Torres-Madronero et al. [22] used a similar instrument to estimate canopy water content and to build a spectral library of maize crops under different nitrogen treatments. Elmer et al. [23] developed a MATLAB-based tool to process the spectral datasets. More information on the ASD Fieldspec spectroradiometer instrument can be found in work by Milton [24], Elmer et al. [23], and the references therein. The datasets can be used to validate and calibrate satellite products and machine learning algorithms to map crop–soil–water conditions at the satellite level, thus improving the accuracy of relevant models.
Figure 1. Map illustrating distribution of surveys in different provinces.

Table 1. Summary of the datasets by province and municipality. The Project column describes the purpose of the data collected; the Temporal resolution column indicates how often data were collected during specific months; the Spatial Resolution column indicates the pixel size; and the Spectral Data column indicates whether spectral radiometer data are available.

<table>
<thead>
<tr>
<th>Province</th>
<th>Municipality</th>
<th>Project</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
<th>Spectral Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Cape</td>
<td>Frances Baard</td>
<td>Barley evapotranspiration</td>
<td>Monthly, August–October 2020</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Free State</td>
<td>Thabo Mofutsanyana</td>
<td>Erosion modelling</td>
<td>Yearly, August 2021 &amp; August 2022</td>
<td>2 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Limpopo</td>
<td>Vhembe/Mopani</td>
<td>Crop disease</td>
<td>Monthly, 2021–2022</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Limpopo</td>
<td>Vhembe</td>
<td>Crop disease</td>
<td>Monthly, January–March 2021 &amp; January–February 2023</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Limpopo</td>
<td>Vhembe</td>
<td>Soil Moisture</td>
<td>Monthly, January–March 2022</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Limpopo</td>
<td>Vhembe</td>
<td>Crop estimate</td>
<td>Daily, March-2020</td>
<td>2–8 cm</td>
<td>No</td>
</tr>
<tr>
<td>Limpopo</td>
<td>Vhembe</td>
<td>4IR in farming</td>
<td>Monthly, August–November 2022</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Eastern Cape</td>
<td>Chris-Hani</td>
<td>Crop estimate</td>
<td>Daily, February 2022</td>
<td>2–8 cm</td>
<td>No</td>
</tr>
<tr>
<td>Eastern Cape</td>
<td>OR-Thambo</td>
<td>Bush Encroachment</td>
<td>Monthly, October–December 2022 &amp; March 2023</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Eastern Cape</td>
<td>Amathole</td>
<td>Bush Encroachment</td>
<td>Monthly, October–December 2022 &amp; March 2023</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Free State</td>
<td>Thabo Mofutsanyana</td>
<td>4IR in farming</td>
<td>Monthly, July–October 2021</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Gauteng</td>
<td>Tshwane Municipality</td>
<td>4IR in farming</td>
<td>Monthly, February–May 2022</td>
<td>2–8 cm</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2. Data Description

The datasets consist of UAV ultra-high spatial resolution images and spectra collected at different locations in South Africa. The UAV surface reflectance images are stored as raster Tag Image File Format (TIFF). The images were projected to EPSG:32736-WGS 84/UTM zone 36S. The datasets contain blue, green, red, near-infrared, and red-Edge bands at the specified bandwidths (Table 2). The spectral data are stored in table format, where the columns indicate the spectra of samples and the rows consist of sample spectra.
from 340 nm to 2150 nm. For comparison purposes, Table 2 also lists band properties of Sentinel-2 datasets. Not all the bands are centered at the common wavelengths and Sentinel-2 has a much narrower bandwidth compared to the RedEdge-MX sensor used in this study.

Table 2. Comparison of band properties of the RedEdge-MX sensor and Sentinel-2 multispectral imagery.

<table>
<thead>
<tr>
<th>Band Name</th>
<th>RedEdge-MX Sensor</th>
<th>Sentinel-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Center Wavelength (nm)</td>
<td>Bandwidth (nm)</td>
</tr>
<tr>
<td>Blue</td>
<td>475</td>
<td>20</td>
</tr>
<tr>
<td>Green</td>
<td>560</td>
<td>20</td>
</tr>
<tr>
<td>Red</td>
<td>668</td>
<td>10</td>
</tr>
<tr>
<td>Near Infrared</td>
<td>840</td>
<td>40</td>
</tr>
<tr>
<td>Red-Edge</td>
<td>717</td>
<td>10</td>
</tr>
</tbody>
</table>

3. Methods
3.1. Unmanned Aerial Vehicle (UAV)
3.1.1. Data Collection
The UAV multi-rotor DJI-Matrice 600 (Figure 2) was used to collect near-ground-based imagery for different crops and natural resources. The altitude was varied from 50 m to 100 m, resulting in 2 cm to 8 cm spatial resolution imagery. A five-band sensor was used to capture imagery (Table 1).

Figure 2. Multi-rotor DJI-Matrice 600 and the MicaSense RedEdge-MX sensor used in the study. “Light and directional sensor” refers to the Downwelling Light Sensor (DLS).

3.1.2. Pre-Processing
The sensor has bands comparable to those of publicly available satellite data such as Sentinel-2 data. The UAV images are radiometrically corrected by using the Calibration Reflectance Panel (CRP) and the sensor measurements of irradiance from the Downwelling Light Sensor (DLS). The CRP contains unique laboratory-calibrated reflectance values that are used together with DLS data to generate surface reflectance. Within PIX4D mapper software [16], the following factors are taken into account: the sensor black level, the
sensitivity of the sensor, sensor gain and exposure settings, and lens vignette effects to generate atmospherically corrected surface reflectance images. The values of the surface reflectance range from 0 to 1.

3.1.3. Example of Results

Figure 3 compares a Sentinel-2 image to the UAV image of a wheat farm in Free State Province. The Sentinel image has a 10 m spatial resolution, while the UAV image has a 2 cm spatial resolution. The advantage of the Sentinel-2 data is that it has more spectral information (Figure 2) and high temporal resolution, which is fixed at about 6 to 10 days. The images can be affected by clouds. The UAV platform is more flexible, it is less affected by clouds and images can be acquired at any time as required by the farmer. Sentinel-2 A + B satellites belong to the European Space Agency [6] and the data are provided free to the public.

Figure 3. Comparison of Sentinel-2 imagery at 10 m spatial resolution (13 September 2021) and UAV imagery at 2 cm spatial resolution (10 September 2021). The yellow and red squares indicate the position of the zoomed inserts that are displayed on the right-side from top to bottom respectively.

Figure 4 compares the Normalized Difference Vegetation Index (NDVI), which is commonly used to assess crop health status [25,26]. It is computed using Equation (1) [27]:

\[
NDVI = \frac{NIR - RED}{NIR + RED},
\]

where \(NIR\) refers to the near-infrared band and \(RED\) refers to the red bands (see Table 2). The NDVI values range from \(-1\) to \(1\), a classification scheme adapted from Mangewa et al. [28]. It is clear from the UAV imagery that more detailed information can be extracted, and crop conditions can be assessed more accurately. The differences in surface reflectance values or derived NDVI values may be due to the Sentinel-2 surface reflectance values derived using Sen2Cor software version 2.9 [29], while UAV imaging used Pix4D software.
Figure 4. NDVI images of Sentinel-2 and the UAV as derived from Figure 3 indicating crop condition. The purple square indicates the position of the zoomed insert on the right-hand side.

3.2. Spectroradiometer Data
3.2.1. Data Collection

An ASD Fieldspec spectroradiometer (Malvern Panalytical Ltd., Malvern, UK) was used to collect field spectra across the 400 to 1000 nm wavelength range at a spectral resolution of 3 nm. Spectral reflectance measurements represent the amount of energy that is reflected by the surface at specific wavelengths [30]. Generally, the field measurement procedure involves (1) calibrating the instrument using the flux reflected from the reference surface (e.g., Spectralon panel) and (2) measuring the flux from the subject (e.g., wheat crop) [31]. A repetition rate was set to five measurements per point to improve the signal-to-noise ratio. Measurements were geo-located with a Global Positioning System (GPS) device connected to the ASD instrument.

3.2.2. Pre-Processing

The raw spectral data were processed to reflectance using the ViewSpec™ Pro 4.02 software provided by the Malvern Panalytical Ltd., Malvern, UK company. The reflectance data were passed through a Savitzky–Golay filter based on a simplified least-squares convolution method [32]. This method can remove measurements that are affected by random noise, clouds, or poor atmospheric conditions [33]. Equation (2) describes the Savitzky–Golay filter algorithm. $Y^*_j$ represents filtered values, $C_i$ is the coefficient of the $i$-th spectral value filter, $Y_j$ represents the original spectrum values, and $N$ is the number of convoluting integers equal to the smoothing window size $(2m + 1)$. 

$$Y^*_j = \sum_{i=0}^{N} C_i Y_{j+i}$$
The smoothed reflectance data were corrected using Multiplicative Scatter Correction (MSC) according to Martens et al. [34], Isaksson and Næs [35], and Geladi et al. [36]. MSC removes additive and/or multiplicative effects on the spectral reflectance data. This method can compensate for differences in the baseline and trend and relatively maintains the original shape of the spectra, making it easy to interpret the transformed data [37]. Each individual spectrum is regressed on the set-mean-spectrum $L_m$ and the linear regression coefficients are estimated according to Equation (3):

$$L_i = a_i + b_i L_m + e_i,$$

where $a_i$ is the intercept estimating the proportional additive effects and the slope $b_i$ estimates the multiplicative effects or rotation away from the line of unity. The errors $e_i$ are attributed to chemical information [38]. Finally, MSC is corrected according to Equation (4):

$$M_i = (L_i - a_i) / b_i.$$

The first derivative $D_{\lambda + \Delta \lambda / 2}$ can be calculated using Equation (5) by taking the difference in reflectance between two closely spaced wavelengths $\Delta \lambda$. The second derivative $D_{\lambda}$ as defined in Equation (6) can be derived from Equation (5):

$$D_{\lambda + \Delta \lambda / 2} = \frac{A_{\lambda + \Delta \lambda / 2} - A_{\lambda}}{\Delta \lambda},$$

$$D_{\lambda} = \frac{A_{\lambda - \Delta \lambda / 2} - 2A_{\lambda} + A_{\lambda + \Delta \lambda / 2}}{\Delta \lambda^2}.$$

For a given spectrum $f(\lambda)$, the Continuous Wavelet Transform (CWT) can be computed using Equation (7) [39]:

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(\lambda) \psi\left(\frac{\lambda - b}{a}\right) d\lambda,$$

where $\lambda$ is the wavelength, $(\lambda)\psi$ is a mother wavelet, $a$ is the scaling factor, and $b$ is the shifting factor. To identify the characteristic wavelengths, the first derivative, second derivative, and CWT of the MSC-corrected reflectance curves can be used. All these methods have been used in the past to extract or characterize spectral reflectance. The CWT method is thought to reduce noise and enhance spectral features [40]. Xu et al. [41] noted an improvement in the classification of mangrove species when applying CWT to hyperspectral data.

### 3.2.3. Example of Results

Figures 5–10 illustrate examples of wheat spectra, with different models applied to the datasets as described above. Such datasets can complement satellite or UAV data at the farm level for the purposes of understanding the variability of nutrients such as nitrogen, chlorophyll, crop stress, and water content. Table 1 describes the availability of spectral data for different projects. Detailed hyperspectral applications are provided by Thenkabail et al. [42] for further reference.
Figure 5. Example of the spectral profile of wheat crops; the spectra depict the wavelength range from 400 to 1000 nm.

Figure 6. Smoothed spectra using Savitzky–Golay filter from Equation (2).
Figure 7. MSC-derived spectral values using Equations (3) and (4).

Figure 8. First-derivative-derived spectral values using Equation (5).
3.2.4. Scientific Importance and Use of UAVs for Precision Agriculture and Natural Resources

The database generated provides a rich dataset for satellite sensor calibration and a reference for the next generation of satellite sensors. South Africa is a developing country with huge potential in space science. The next generation of satellites developed by institutions such as the South African Space Agency (SANSA), Cape Peninsula University...
of Technology, and associated institutions will rely on ground-based validation datasets to design and calibrate appropriate sensors that directly/indirectly measure crop stress, water, nutrients, disease, etc. UAVs and spectral field data of different crops captured at spatial and temporal resolutions can contribute towards the next generation of sensors. UAV data have been shown to possess useful applications in precision agriculture. Examples of these include applications of UAVs for sugarcane crops [43], as well as precision agriculture applications [43–45] and their associated challenges [46].

Different spectral libraries of different material exist. Examples include a soil spectral library [47]; an integrated open mineral spectral library developed by Xie et al. [48] (Rock Spectral Library, RockSL); the wetlands spectral library [49,50]; the ECOSTRESS spectral library [51]; and a maize spectral library under nitrogen deficiency developed by Torres-Madronero et al. [22], to name a few. Most of these efforts have focused on measuring spectral signatures of different materials based on single measurements, i.e., most spectral libraries lack temporal measurement components. For applications such as those concerning minerals and rocks, a multi-temporal spectral library is not necessarily needed. However, for vegetation and crops, it is necessary as the spectra can be influenced by growth stage and biochemical properties (moisture content, chlorophyll, nitrogen, sugar) [52–54]. Therefore, the spectral datasets provided offer an opportunity to evaluate crops at different crop stages.

3.2.5. Challenges in Developing UAV and Spectral Databases in South Africa

The main challenge is the funding required to have multiple UAV systems that can cater for the demands of the agricultural sector, Department of Environmental Affairs, and research institutions. The government of South Africa has placed high regulations around the use of UAV systems, especially for commercial purposes, and such regulations have tended to slow down the uptake of this technology. Training of pilots and funding for fieldwork also makes it difficult to consistently collect UAV imagery and spectral data. UAV systems have limited flight time and require regular maintenance. This makes it impossible to cover large areas during surveys. A web-based data repository should be implemented by the ARC-NRE to enable easy access of UAV imagery. This should be done in line with institutional policy on data sharing.

3.2.6. Data Availability

Datasets are available on request:
ARC-Natural Resources and Engineering
Tel.: +27-(0)12-310-2500
Fax: +27-(0)12-323-1157
E-mail: munghezuluc@arc.agric.za

Physical address: 600 Belvedere Street, Arcadia, Pretoria, South Africa
Postal address: Private Bag X79, Pretoria, 0001
GPS coordinates: 25°44’19.4” S, 28°12’26.4” E

4. Conclusions and Recommendations

This paper presented UAV and spectral datasets collected from different projects performed by the ARC-NRE. The projects cover different provinces within South Africa. The datasets are collected from farm areas and some of the datasets concern natural resources. The data can be used to benefit wider scientific communities and encourage collaboration. The UAV data collected by the ARC-NRE for various natural resources form a basis for long-term monitoring and detecting changes in natural resources. By comparing UAV images collected at different time points, researchers can quantify the rate and extent of these changes, identify hotspots, and support conservation and land management efforts. UAVs capture detailed imagery that helps in assessing vegetation health, biomass estimation, and species composition as demonstrated by the vegetation
indices generated in this paper. These data aid in crop monitoring, weed detection, and improving farm management.

There is a greater need to use the data to support smallholders and emerging farmers across South Africa. However, the cost of data acquisition needs to be reduced by allocating more funding to similar projects. A central database should be developed to encourage data sharing.

Author Contributions: Conceptualization, methodology, analysis, and original draft preparation, C.M.; writing—review and editing, Z.M.-M.; writing—review and editing, funding acquisition, S.S. and G.C.; fieldwork operations and editing, P.E.R. and E.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agricultural Research Council-Natural Resources and Engineering (ARC-NRE), Department of Science and Innovation, Council for Scientific and Industrial Research, grant number P07000198; the National Research Foundation (NRF-Thuthuka, grant number TTK200221506319); and the Department of Agriculture, Land Reform and Rural Development (DALRRD).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Available on request to munghemezuluc@arc.agric.za.

Acknowledgments: We would also like to thank the following people for participating during the field campaigns: Wonga Masiza; Pitso Walter Khoboko; Nombuso Parkies; Colette De Villiers; Vuwani Makuya; Shaun Muerheid; Idani Gundula; Tsimangadzo Rasifudi; Siyabonga Gasa; Siboniso Nkambule; Anneka Koaanni; Michael Kidson; Siphokazi Gcayi; Walter Khoboko; Adolph Nyamugama; Khaled Abutaleb; Basani Nkuna; Walter Khoboko; Richard Tswai; Nombuso Parkies; Mohamed Elbasit; Johan Van Biljon; and Solomon W. Newete.

Conflicts of Interest: The authors declare no conflict of interest.

References


27. Honda, R.; McCabe, M. High-Resolution NDVI from Planet’s Earth Observing Nano-Satellites: A New Data Source for Precision Agriculture. *Remote Sens.* 2016, 8, 768. [CrossRef]


33. Isaksson, T.; Næs, T. The Effect of Multispectral Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Appl. Spectrosc.* 1988, 42, 1273–1284. [CrossRef]


36. Dhanola, M.S.; Lister, S.J.; Sanderson, R.; Barnes, R.J. The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *J. Near Infrared Spectrosc.* 1994, 2, 43–47. [CrossRef]


