Data in Brief 46 (2023) 108786

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Satellite imagery dataset of manure application on pasture fields

Oscar D. Pedrayes, Rubén Usamentiaga*

Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, Gijón, Asturias 33204, Spain

ARTICLE INFO

Article history:
Received 9 August 2022
Revised 2 November 2022
Accepted 24 November 2022
Available online 29 November 2022

Keywords:
Fertilizer
Land
Slurry
Crops
Semantic segmentation
Classification
Precision agriculture

ABSTRACT

Applying manure to pasture fields is a very common method of fertilization. However, rainfall can cause the manure to leach into water bodies near the field, contaminating the water and damaging the environment and the animals living in it, ultimately affecting human life. This paper presents a dataset consisting of images of 30 plots after manure application, verified by on-site investigations. This involved visiting 38 different plots, of which 8 were discarded because they were not suitable, either because of their small size, the lack of a specific manure application date, or the images being too cloudy in that period. The imagery is collected through Google Earth Engine using the satellite Sentinel-2, which offers 13 hyperspectral bands in the range of ultraviolet and near-infrared wavelengths including the visible spectrum. From these 13 bands, the most common hyperspectral indices in the literature for precision agriculture are calculated and added into the images as channels. 51 hyperspectral indices are calculated, summing up to a total of 64 channels per image when adding the raw bands from Sentinel-2. No normalization has been performed on any of the channels. The data can be used for further research of automatic classification of manure application to control its use and prevent contamination.

* Corresponding author.
E-mail addresses: UO251056@uniovi.es (O.D. Pedrayes), rusamentiaga@uniovi.es (R. Usamentiaga).

https://doi.org/10.1016/j.dib.2022.108786

Specifications Table

Subject	Agronomy and Crop Science
Specific subject area	Remote sensing for precision agriculture to detect and classify recently manured pasture fields.
Type of data	Image
How the data were acquired	All data is acquired through Google Earth Engine. Plots are manually selected in Google Earth Engine after an on-site investigation. Images are downloaded from the satellite Sentinel-2 using Google Earth Engine. Finally, cloudy images are filtered out manually.
Data format	Raw Filtered Processed
Description of data collection	The regions of interest with the considered plots in the images are located after careful on-site inspection and verification. When a manured field is found, photographs are taken as validation and the location is indicated by GPS. Then, from Google Earth Engine, the appropriate region is manually selected, and the corresponding Sentinel-2 images are downloaded from the date on which the plot was manured, or the closest possible later date.
Data source location	City/Town/Region: Northern region of Spain Country: Spain Latitude and longitude: latitude around [43.55, 43.38], and longitude around [−5.50, −4.10]. All images are obtained from the satellite Sentinel-2.
Data accessibility	Repository name: Mendeley Data Data identification number: https://doi.org/10.17632/fbvvvf55kp.1 Direct URL to data: https://data.mendeley.com/datasets/fbvvvf55kp

Value of the Data

- The development of such a dataset is costly and time-consuming, as on-site investigations are necessary to verify manure application and to accurately select the plot. In addition, clouds and other problems, such as plots that are too small, must be filtered out.
- This dataset can be used to train machine learning models to automatically detect manured fields to analyze illegal fertilization or hot spots. This provides an opportunity for further research on this topic.
- Each plot has multiple images from different dates from before and after manure application. This offers the opportunity to investigate classification methods that benefit from temporal analysis. The differences in terrain depending on its date can be considerable, which adds a substantial amount of information to the status of the plot.
- The imagery contains the most relevant hyperspectral indices in the literature for precision agriculture and provides all 13 Sentinel-2 bands from which more hyperspectral indices can be created if needed.
- There is no other dataset of this type in the literature for this particular problem. Moreover, even if other datasets were created, this data would still be useful, as it belongs to a particular region and crop type which could be used to complete other datasets or to validate results.

1. Data Description

The dataset consists of three folders: the "src" folder, where all the code to generate the dataset is stored; the "groundtruth" folder, which contains an image mask for each plot; and the "imagery" folder which contains images with the satellite imagery raw bands and the calculated hyperspectral indices. The ground truth images are in ".png" format and follow a color code:

- White (255, 255, 255): Plot of interest
- Black (0,0,0): Other

The "imagery" folder contains a folder for each plot. Each plot folder contains another two folders, one for the images from before the application of manure and another one for the images from after manure application. Every image is in ".tif" format and has 64 channels. The order of the channels and how to calculate them can be found in the "Experimental design, materials and methods" section.

All the plots obtained for this dataset are pastures. This is because pasture is the predominant type of crop in this area of northern Spain. In most cases the grass is mowed prior to manure application. Although in some of the plots the manure is applied directly on the plowed land. This could prevent the trained models from confusing plowed lands and manure. Images of the plots have an area of about 1700 $\times$ 1700 m, although the plots inside the images are smaller. A total of 38 plots are studied.

Table 1 summarizes every plot of interest in the dataset, showing the date of manure application, area in square meters, number of available images for each plot from before and after manure application, its suitability for further studies, and its geographical coordinates. The plot identifier is composed of "P-" plus the abbreviation (using only the consonants) of the locality in which the plot is located. The area of the plots is calculated after generating the ground truth mask, where each Sentinel-2 pixel counts as 100 m^2. The suitability is assessed after studying the Sentinel-2 images of the plot in question. For example, if the region is extremely small, it is discarded.

Table 1
Dataset summary.

Plot	Date (YYYY/MM/dd)	Area (m^2)	Available images (Before/After)	Suitable (Yes/No)	Geographical Coordinates (Long/Lat)	
P-BLD	2022/05/26	8900	2/1	**Yes**	−4.2018	43.3973
P-BLLT1	2022/05/16	21,200	2/2	**Yes**	−4.0840	43.4309
P-BLLT2	2022/05/26	3300	2/1	**Yes**	−4.0840	43.4310
P-Cardana	2022/02/24	6500	8/9	**Yes**	8.6580	45.8592
P-CBRCS1	2022/05/26	6700	2/1	**Yes**	−4.2005	43.3897
P-CBRCS2	2022/05/26	6400	2/1	**Yes**	−4.2048	43.3875
P-CLGT	2022/05/16	17,200	3/2	**Yes**	−4.1096	43.3987
P-CLMBRS	2022/05/26	4300	3/1	**Yes**	−4.5447	43.3804
P-CMNTR	2022/05/16	2600	1/2	**Yes**	−4.1470	43.4001
P-DR.	2022/03/21	2500	1/5	**Yes**	−4.1424	43.3967
P-FNFR	2022/05/16	10,100	2/2	**Yes**	−4.2657	43.3880
P-GLS	2022/04/30	7800	2/-	**No** (Clouds)	−4.1452	43.3996
P-LLT	2022/05/03	9600	2/1	**Yes**	−4.1515	43.4001
P-LNDRS1	2022/05/16	3200	2/2	**Yes**	−4.2510	43.3880
P-LNDRS2	2022/05/16	5400	2/2	**Yes**	−4.2503	43.3880
P-LNDRS3	2022/05/16	8500	2/2	**Yes**	−4.2497	43.3872
P-LNDRS4	2022/05/16	9100	2/2	**Yes**	−4.2467	43.3877
P-LNDRS5	–	5100	2/2	**No** (application date unclear)	−4.2435	43.3864
P-MT	2022/05/04	19,900	2/1	**Yes**	−4.1536	43.3980
P-NMS	2022/02/10	5500	2/1	**Yes**	−4.1490	43.4003

(*continued on next page*)

Table 1 (*continued*)

Plot	Date (YYYY/MM/dd)	Area (m^2)	Available images (Before/After)	Suitable (Yes/No)	Geographical Coordinates (Long/Lat)	
P-PQN	2022/02/27	5300	-/2	**No**. (Clouds)	−4.1495	43.3991
P-PSG	2022/04/06	5400	3/2	**No**. (Too narrow, partly fertilized)	−4.1411	43.3970
P-QNTLS1	–	13,600	7/3	**No** (application date unclear)	−5.5830	43.5463
P-QNTLS2	2022/05/16	8500	7/3	**Yes**	−5.5840	43.5458
P-SNTLLN	2022/03/17	14,200	2/4	**Yes**	−4.1170	43.3935
P-SNVCNT1	2022/05/16	6700	2/2	**Yes**	−4.4048	43.3939
P-SNVCNT2	2022/05/16	29,200	2/2	**Yes**	−4.4001	43.3945
P-STBN	2022/05/04	11,300	3/1	**Yes**	−4.1366	43.3960
P-TGL1	–	28,000	2/1	**No** (application date unclear)	−4.0695	43.4216
P-TGL2	2022/05/16	12,300	2/2	**Yes**	−4.0701	43.4276
P-TMSN	2022/02/10	4700	–	**No** (Clouds)	−4.1519	43.3996
P-TNNS1	2022/05/26	19,500	2/1	**Yes**	−4.1871	43.3999
P-TNNS2	2022/05/06	11,100	1/2	**Yes**	−4.1918	43.3987
P-TPRN	2022/04/06	1800	3/3	**No**. (Too narrow)	−4.1390	43.3965
P-VG1	2022/04/09	12,200	3/6	**Yes**	−5.4866	43.4699
P-VG2	2022/04/13	4900	4/5	**Yes**	−5.4801	43.4693
P-VLDMR	2022/02/07	17,500	2/2	**Yes**	−4.1561	43.4056
P-VNS	2022/04/23	16,600	3/2	**Yes**	−4.1504	43.4042

The complete dataset consists of 31.48 ha for the plots of interest. Each pixel is 0.01 ha.

2. Experimental Design, Materials and Methods

The first indications of a newly manured plot are given by people living in the area or by Sentinel-2 imagery surveys. To label the plots, first, an on-site investigation is carried out to confirm that the plot has been fertilized with manure and to observe the real dimensions of the fertilization in the plot. Then, using Google Earth Engine, the plot is annotated according to the observed dimensions. An example of this process is shown in Fig. 1.

Google Earth Engine is also used to download the imagery with the script “download_imagery.js”. The plots are then exported as KML files and used to generate ground truth masks by taking advantage of the georeferenced Sentinel-2 imagery, as shown in Fig. 2. The script used to generate the ground truth masks from the KML files is called “generate_groundtruth.py”

To generate the images of the dataset, the first 13 channels are obtained directly from the 13 bands of the Sentinel-2 images. Sentinel-2 has two satellites in its orbit (Sentinel-2A and Sentinel2B), where each one has an orbit of 10 days. Their orbits are at the greatest distance from each other, which is why the acquisition time of the images for the same region is around 5 days. Table 2 shows the wavelengths and bandwidths for each band in µm for S-2A and S-2B separately.

The remaining 51 channels of the images from the dataset are hyperspectral indices which are calculated as different combinations of the 13 Sentinel-2 raw bands. To obtain these hyperspectral indices, the general literature of precision agriculture for fertilizers using satellite imagery has been studied [1–9] and the most relevant hyperspectral indices have been obtained. The script necessary to generate the hyperspectral indices is called “calculate_indices.py”. Table 3 shows how to calculate each hyperspectral index and its channel number in the images.

Fig. 1. On-site investigation (left). Plot annotated in Google Earth Engine (right).

Fig. 2. Sentinel-2 georeferenced image (left). Generated ground truth (right).

To get an idea of the final images, an example of a visualization of one of the plots is shown. Fig. 3 shows at the left the total area of the image in RGB, and at the right, an enlarged version. Fig. 4 shows each of the 64 channels in a black and white color scale. To better visualize the plot, an enlarged version is shown in Fig. 5.

Table 2
Sentinel-2 bands (Sentinel2A and Sentinel2B).

#	Band	Central Wavelength (μm)		Bandwidth (μm)		Spatial resolution (m)
		S-2A	S-2B	S-2A	S-2B	
0	B01 Coastal aerosol	0.4427	0.4422	0.021	0.021	60
1	B02 Blue	0.4924	0.4921	0.066	0.066	10
2	B03 Green	0.5598	0.5590	0.036	0.036	10
3	B04 Red	0.6646	0.6649	0.031	0.031	10
4	B05 VRE	0.7041	0.7038	0.015	0.016	20
5	B06 VRE	0.7405	0.7391	0.015	0.015	20
6	B07 VRE	0.7828	0.7797	0.020	0.020	20
7	B08 NIR	0.8328	0.8329	0.106	0.106	10
8	B8A Narrow Nir	0.8647	0.8640	0.021	0.022	20
9	B09 Water vapor	0.9451	0.9432	0.020	0.021	60
10	B10 SWIR Cirrus	1.3735	1.3769	0.031	0.030	60
11	B11 WIR	1.6137	1.6104	0.091	0.094	20
12	B12 SWIR	2.2024	2.1857	0.175	0.185	20

Table 3
Hyperspectral indices.

#	Abb.	Name	Description
13	**NDVI**	Normalized Difference Vegetation Index	$\frac{B08 - B04}{B08 + B04}$
14	**NSNDVI**	NIR-SWIR Normalized Difference Vegetation Index	$\frac{B11 - B07}{B11 + B07}$
15	**SDI**	Swir Difference Index	$B08 - B12$
16	**GNDVI**	Green Normalized Difference Vegetation Index	$\frac{B08 - B03}{B08 + B03}$
17	**SAVI**	Soil Adjusted Vegetation Index	$\frac{B08 - B04}{B08 + B04 + 0.428} * 1.428$
18	**OSAVI**	Optimized Soil Adjusted Vegetation Index	$(1 + 0.16) * \frac{B08 - B04}{B08 + B04 + 0.16}$
19	**EOMI1**	Exogenous Organic Matter Index 1	$\frac{B11 - B8A}{B11 + B8A}$
20	**EOMI2**	Exogenous Organic Matter Index 2	$\frac{B12 - B04}{B12 + B04}$
21	**EOMI3**	Exogenous Organic Matter Index 3	$\frac{(B11 - B8A) + (B12 + B04)}{B11 + B8A + B12 + B04}$
22	**EOMI4**	Exogenous Organic Matter Index 4	$\frac{B11 - B04}{B11 + B04}$
23	**BNR2**	Normalized Burn Ratio 2	$\frac{B11 - B12}{B11 + B12}$
24	**RVI**	Ratio Vegetation Index	$\frac{B08}{B04}$
25	**DVI**	Difference Vegetation Index	$B08 - B04$
26	**RENDVI1**	Red Edge Normalized Difference Vegetation Index	$\frac{B05 - B04}{B05 + B04}$
27	**RENDVI2**	Red Edge Normalized Difference Vegetation Index	Same as RENDVI1, but uses B06 instead of B05
28	**RENDVI3**	Red Edge Normalized Difference Vegetation Index	Same as RENDVI1, but uses B07 instead of B05
29	**CI1**	Chlorophyll Index	$\frac{B08}{B05} - 1$
30	**CI2**	Chlorophyll Index	Same as CI1, but uses B06 instead of B05
31	**CI3**	Chlorophyll Index	Same as CI1, but uses B07 instead of B05
32	**NDRE**	Normalized Difference Red Edge	$\frac{B08 - B05}{B08 + B05}$

(*continued on next page*)

Table 3 (*continued*)

#	Abb.	Name	Description
33	**MCARI**	Modified Chlorophyll Absorption in Reflectance Index	$((B05 - B04) - 0.2 * (B05 - B03)) * \frac{B05}{B04}$
34	**MCARI1**	Modified Chlorophyll Absorption in Reflectance Index 1	$1.2 * (2.5 * (B08 - B04) - 1.3 * (B08 - B03))$
35	**MCARI2**	Modified Chlorophyll Absorption in Reflectance Index 2	$1.5 * \frac{2.5 * (B08 - B04) - 1.3 * (B08 - B03)}{\sqrt{(2 * B08 + 1)^2 - (6 * B08 - 5 * \sqrt{B04}) - 0.5}}$
36	**MTVI1**	Modified Triangular Vegetation Index 1	$1.2 * (1.2 * (B08 - B03) - 2.5 * (B04 - B03))$
37	**MTVI2**	Modified Triangular Vegetation Index 2	$1.5 * \frac{1.2 * (B08 - B03) - 2.5 * (B08 - B03)}{\sqrt{(2 * B08 + 1)^2 - (6 * B08 - 5 * \sqrt{B04}) - 0.5}}$
38	**EVI**	Enhanced Vegetation Index	$\frac{2.5 * (B08 - B04)}{(B08 + 6 * B04 - 7.5 * B02) + 1}$
39	**AVI**	Advanced Vegetation Index	$(B08 * (1 - B04) * (B08 - B04))^{1/3}$
40	**GCI**	Green Coverage Index	$\frac{B09}{B03} - 1$
41	**BSI**	Bare Soil Index	$B11 + B04 + \frac{B08 + B02}{B11 + B04} + B08 + B02$
42	**NBRI**	Normalized Burned Ratio Index	$\frac{B08 - B12}{B08 + B12}$
43	**NDRE1**	Normalized Difference Red Edge	$\frac{B08 - B05}{B08 + B05}$
44	**NDRE2**	Normalized Difference Red Edge	Same as NDRE1, but uses B06 instead of B05
45	**NDRE3**	Normalized Difference Red Edge	Same as NDRE1, but uses B07 instead of B05
46	**MSAVI**	Modified Soil Adjusted Vegetation Index	$\frac{(2.0 * B08 + 1 - \sqrt{((2.0 * B08 + 1.0)^2 - 8 * (B08 - B04))}}{2}$
47	**WDRVI**	Wide Dynamic Range Vegetation Index	$\frac{0.1 * B08 - B04}{0.1 * B08 + B04}$
48	**ARVI1**	Atmospherically Resistant Vegetation Index 1	$\frac{B8A - B04 - 0.069 * (B04 - B02)}{B8A + B04 - 0.069 * (B04 - B02)}$
49	**ARVI2**	Atmospherically Resistant Vegetation Index 2	$-0.18 + 1.17 * \frac{B8 - B4}{B8 + B4}$
50	**TSAVI**	Transformed Soil Adjusted Vegetation Index	$\frac{(0.421 * (B08 - 0.421 * B04 - 0.824))}{(B04 + 0.421 * (B08 - 0.824) + 0.114 * (1 + 0.421)^2)}$
51	**CARI1**	Chlorophyll Absorption Ratio Index 1	$\frac{B05}{B04} * \frac{\lvert \frac{(B05 - B03)}{150} * 670.0 + B04 + (B03 - (\frac{(B05 - B03)}{150} * 550))\rvert}{\sqrt{(B05 - B03) / 150^2 + 1}}$
52	**CARI2**	Chlorophyll Absorption Ratio Index 2	$\frac{\lvert(B05 - B03) / 150 * B04 + B04 + B03 - 0.496 * B03\rvert}{\sqrt{(0.496^2 + 1))} * (B05 / B04)}$
53	**CVI**	Chlorophyll Vegetation Index	$\frac{B08 * B04}{B03^2}$
54	**EVI1**	Enhanced Vegetation Index 1	$\frac{2.5 * (B08 - B04)}{(B08 + 6 * B04 - 7.5 * B02) + 1}$
55	**EVI2**	Enhanced Vegetation Index 2	$2.4 * \frac{B08 - B04}{B08 + B04 + 1}$
56	**EVI3**	Enhanced Vegetation Index 3	$2.5 * \frac{B08 - B04}{B08 + 2.4 * B04 + 1}$
57	**SCI**	Soil Composition Index	$\frac{B11 - B08}{B11 + B08}$
58	**GRNDVI**	Green-Red Normalized Difference Vegetation Index	$\frac{B08 - (B03 + B04)}{B08 + (B03 + B04)}$
59	**GBNDVI**	Green-Blue Normalized Difference Vegetation Index	$\frac{B08 - (B03 + B02)}{B08 + (B03 + B02)}$
60	**GLI**	Green Leaf Index	$\frac{2 * B03 - B04 - B02}{2 * B03 + B04 + B02}$
61	**ATSAVI**	Adjusted Transformed Soil-Adjusted Vegetation Index	$\frac{1.22 * (B08 - 1.22 * B04 - 0.03)}{1.22 * B08 + B04 - 1.22 * 0.03 + 0.08 * (1 + 1.22^2)}$
62	**ALTERATION**	Alteration Index	$\frac{B11}{B12}$
63	**CTVI**	Corrected Transformed Vegetation Index	$\frac{((B04 - B03) / (B04 + B03)) + 0.5}{\lvert \frac{B04 - B03}{B04 + B03} \rvert + 0.5 * \sqrt{\lvert \frac{B04 - B03}{B04 + B03} + 0.5\rvert}}$

Fig. 3. Sentinel-2 image of a manured plot. Total area of the image plot (left). Enlarged plot (right).

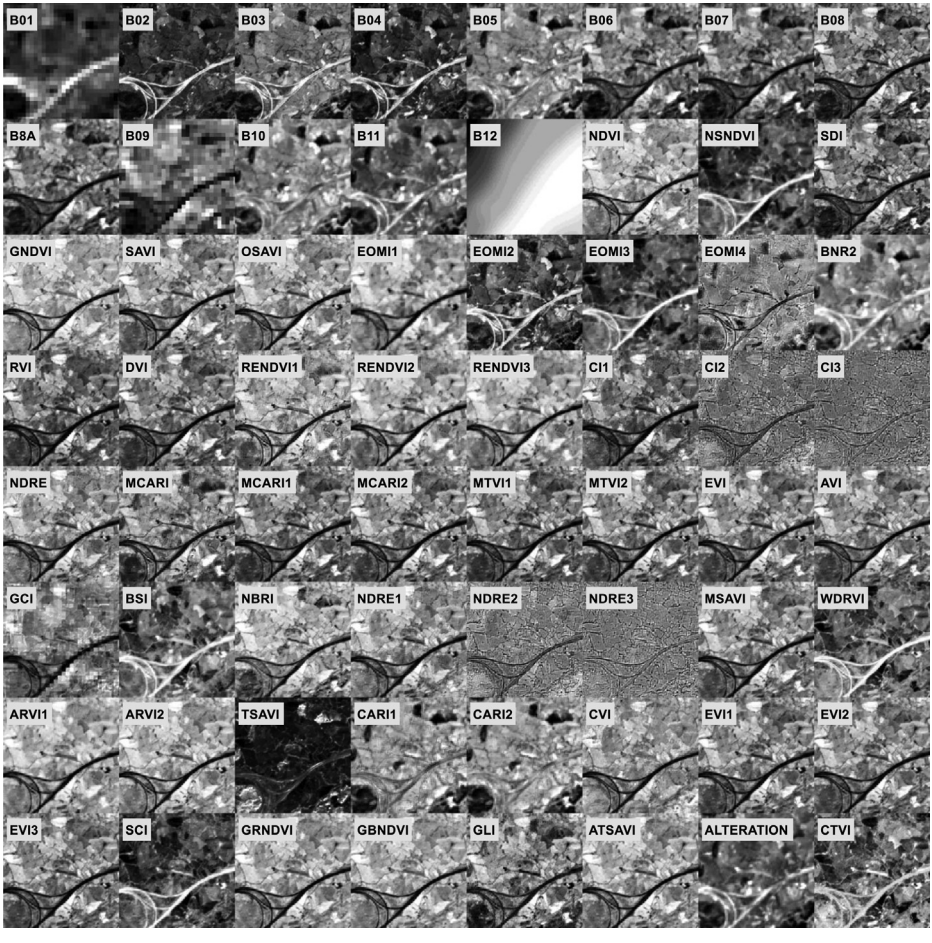

Fig. 4. Example of the 64 channels, including raw band and computed indices.

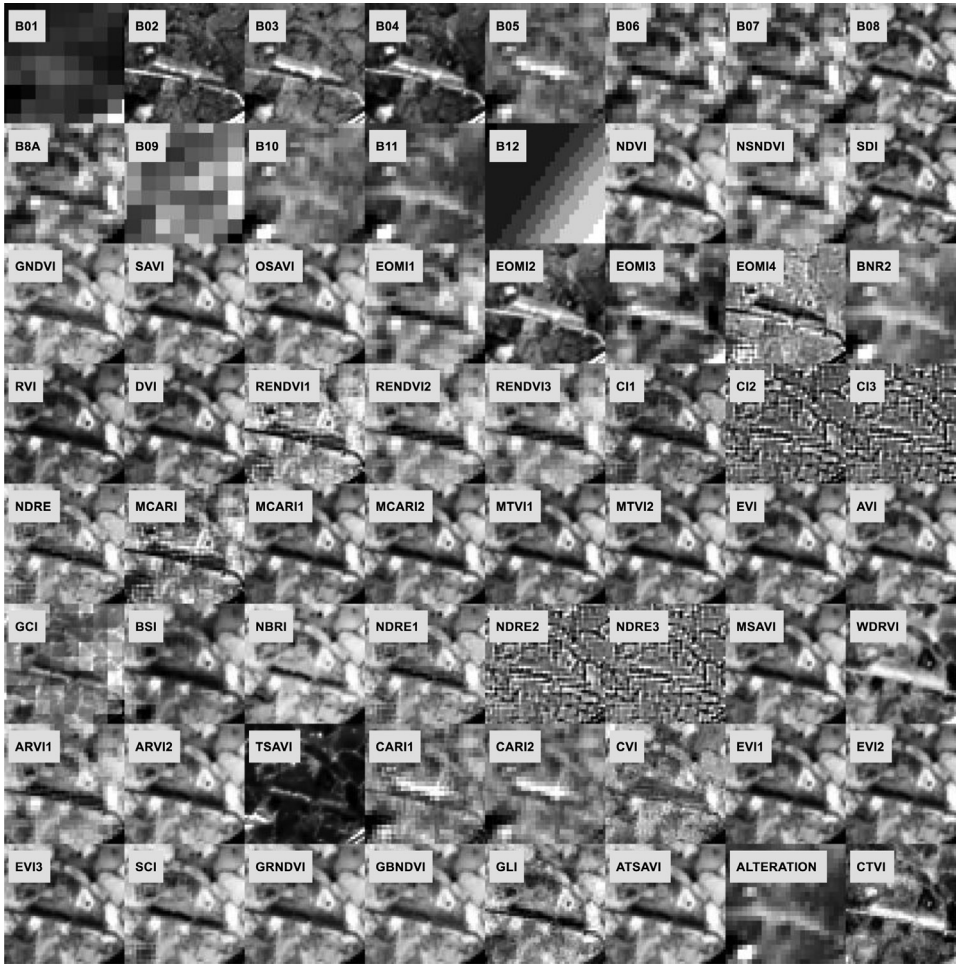

Fig. 5. Example of the 64 channels, including raw bands and computed indices. (Enlarged plot).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Satellite imagery dataset of manure application on pasture fields (Original data) (Mendeley Data).

CRediT Author Statement

Oscar D. Pedrayes: Conceptualization, Methodology, Software, Visualization, Investigation, Writing – original draft; **Rubén Usamentiaga:** Data curation, Supervision, Validation, Writing – review & editing.

Acknowledgments

This work has been partially funded by the project PID2021-124383OB-I00 of the Spanish National Plan for Research, Development and Innovation.

References

[1] Y. Fu, G. Yang, R. Pu, Z. Li, H. Li, X. Xu, C. Zhao, An overview of crop nitrogen status assessment using hyperspectral remote sensing: current status and perspectives, Eur. J. Agron. 124 (2021) 126241.
[2] M. Dodin, H.D. Smith, F. Levavasseur, D. Hadjar, S. Houot, E. Vaudour, Potential of Sentinel-2 satellite images for monitoring green waste compost and manure amendments in temperate cropland, Remote Sens. 13 (9) (2021) 1616.
[3] N. Bagheri, H. Ahmadi, S.K. Alavipanah, M. Omid, Multispectral remote sensing for site-specific nitrogen fertilizer management, Pesqui. Agropecu. Bras. 48 (2013) 1394–1401.
[4] D.G. Lema, O.D. Pedrayes, R. Usamentiaga, D.F. García, Á. Alonso, Cost-performance evaluation of a recognition service of livestock activity using aerial images, Remote Sens. 13 (12) (2021) 2318.
[5] Q. Ma, W. Yu, H. Zhou, The relationship between soil nutrient properties and remote sensing indices in the Phaeozem region of Northeast China, in: Proceedings of the Second International Conference on Computational Intelligence and Natural Computing, 2, IEEE, 2010, pp. 109–112.
[6] Romanko, M. (2017). *Remote sensing in precision agriculture: monitoring plant chlorophyll, and soil ammonia, nitrate, and phosphate in corn and soybean fields* (Doctoral dissertation, Bowling Green State University).
[7] L. Shou, L. Jia, Z. Cui, X. Chen, F. Zhang, Using high-resolution satellite imaging to evaluate nitrogen status of winter wheat, J. Plant Nutr. 30 (10) (2007) 1669–1680.
[8] R.P. Sishodia, R.L. Ray, S.K. Singh, Applications of remote sensing in precision agriculture: a review, Remote Sens. 12 (19) (2020) 3136.
[9] W. Zhu, E.E. Rezaei, H. Nouri, T. Yang, B. Li, H. Gong, Z. Sun, Quick detection of field-scale soil comprehensive attributes via the integration of UAV and sentinel-2B remote sensing data, Remote Sens. 13 (22) (2021) 4716.