Data in Brief 55 (2024) 110739

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

GloCAB cropland field boundary dataset

Joanne V. Hall*, Fernanda Argueta, Louis Giglio

Department of Geographical Sciences, University of Maryland, 2181 Samuel J. LeFrak Hall, 7251 Preinkert Drive, College Park, MD 20742, United States of America

A R T I C L E I N F O

Article history:
Received 18 January 2024
Revised 21 May 2024
Accepted 8 July 2024
Available online 14 July 2024

Dataset link: GloCAB Cropland Field Boundary Dataset (Original data)

Keywords:
Cropland field boundary
Manually-digitized field boundaries
Sentinel-2
Geographical information system
Remote sensing

A B S T R A C T

This dataset consists of 190,832 manually-digitized cropland field boundaries, with associated attributes, within Brazil, Ukraine, United States of America, Canada, and Russia. Specifically, 22 regions of various sizes ($74km^2$ – $38{,}000km^2$) spanning 5 countries were digitized over a range of predominant crop types over different time periods. These field boundaries were drawn over 20 m Sentinel-2 imagery. This field boundary dataset is a byproduct of a larger effort to map cropland burned area (Global Cropland Area Burned: GloCAB product [1]), however, it has several benefits beyond its original intent, including as a training dataset for machine-learning field size analyses, or a dataset to derive cropland field characteristics across different predominant crop types and geographies.

Specifications Table

Subject	Earth and Planetary Science
Specific subject area	Geographical Information System
Data format	Raw Vector data (Shapefiles)
Type of data	Vectors
Data collection	The cropland field boundary vector/feature data were manually digitized using ArcGIS over 22 regions within Brazil, Ukraine, United States of America, Canada,

(*continued on next page*)

* Corresponding author.
E-mail address: jhall1@umd.edu (J.V. Hall).
Social media: @JoanneVHall (J.V. Hall)

https://doi.org/10.1016/j.dib.2024.110739

	and Russia (Figure 1). Sentinel-2 (20 m spatial resolution) imagery were used as the primary base layer for digitization. Specifically, the SWIR band combination (Band 12, 8A, 4) was used as the larger study [1] focused on identifying if the field burned or not during the temporal window. Each field boundary was drawn by an analyst over 20 m Sentinel-2 imagery and quality-checked by a senior team member. In instances where the fields were too small to identify, the analyst digitized a larger boundary encompassing several small fields and designated the polygon with a “NoArea” flag in the attribute table. Each region was assigned a specific mapping date range (between 2016 and 2020; see Table 1 for details).
Data source location	The Sentinel-2 imagery was downloaded from the Sentinel EO-browser (https://apps.sentinel-hub.com/eo-browser/).
Data accessibility	Repository name: Zenodo Direct URL to data: https://doi.org/10.5281/zenodo.10479122
Related Research Article	Hall, J. V., Argueta, F., Zubkova, M., Chen, Y., Randerson, J. T., & Giglio, L. (2024). GloCAB: global cropland burned area from mid-2002 to 2020. *Earth System Science Data, 16*(2), 867–885.

1. Value of the Data

- Remotely-sensed field boundary data are often used in a variety of studies including food security and socio-economic analyses (e.g., [2]). Specifically, these data can provide insights into cropland field characteristics (e.g., density, shapes, sizes, and compactness) and spatial distributions of farmland at various scales. Furthermore, the spatial and contextual information extracted from field boundary features can also be used alongside spectral information to improve cropland classification analyses [e.g., 3].
- These data can benefit the remote sensing, computer science, geospatial, and machine-learning communities focused on agricultural mapping. Cropland field boundary training and validation/reference data are always in high demand in these fields (e.g., [4,5]). Agricultural landscapes are heterogeneous and farm sizes and shapes are influenced by several factors including, crop types, mechanized versus subsidence farming, different geographies, socio-economic factors, and local population density.
- This dataset provides users with a large collection of vectorized field boundaries across different agricultural landscapes. These data can be used to study field-level characteristics across predominant crop types (winter wheat, spring wheat, rice, maize, and sugarcane) within Brazil, Ukraine, United States of America, Canada, and Russia.

2. Background

This newly generated field boundary dataset emerged as a byproduct of a distinct study centered on mapping global cropland burned area (Global Cropland Area Burned; GloCAB; Hall et al., 2024). The utility of these field boundary features extends well beyond the original purpose of GloCAB. Consequently, a separate publication focused on the feature data will make these field boundaries accessible to a broader audience, including their potential use as machine-learning training data.

The GloCAB dataset offers global, monthly information on cropland burned area at a 0.25-degree resolution spanning from 2003 to 2020 (Hall et al., 2024). Given the unique characteristics of cropland fires, a specialized methodology for mapping burned area was employed to enhance the accuracy of the assessments. As part of the broader research initiative, a total of 190,832 fields were digitally delineated and categorized across 22 regions worldwide. These designated burned area reference regions played a crucial role in the GloCAB analysis, contributing to the development of scaling factors that were subsequently applied to MODIS active fire points.

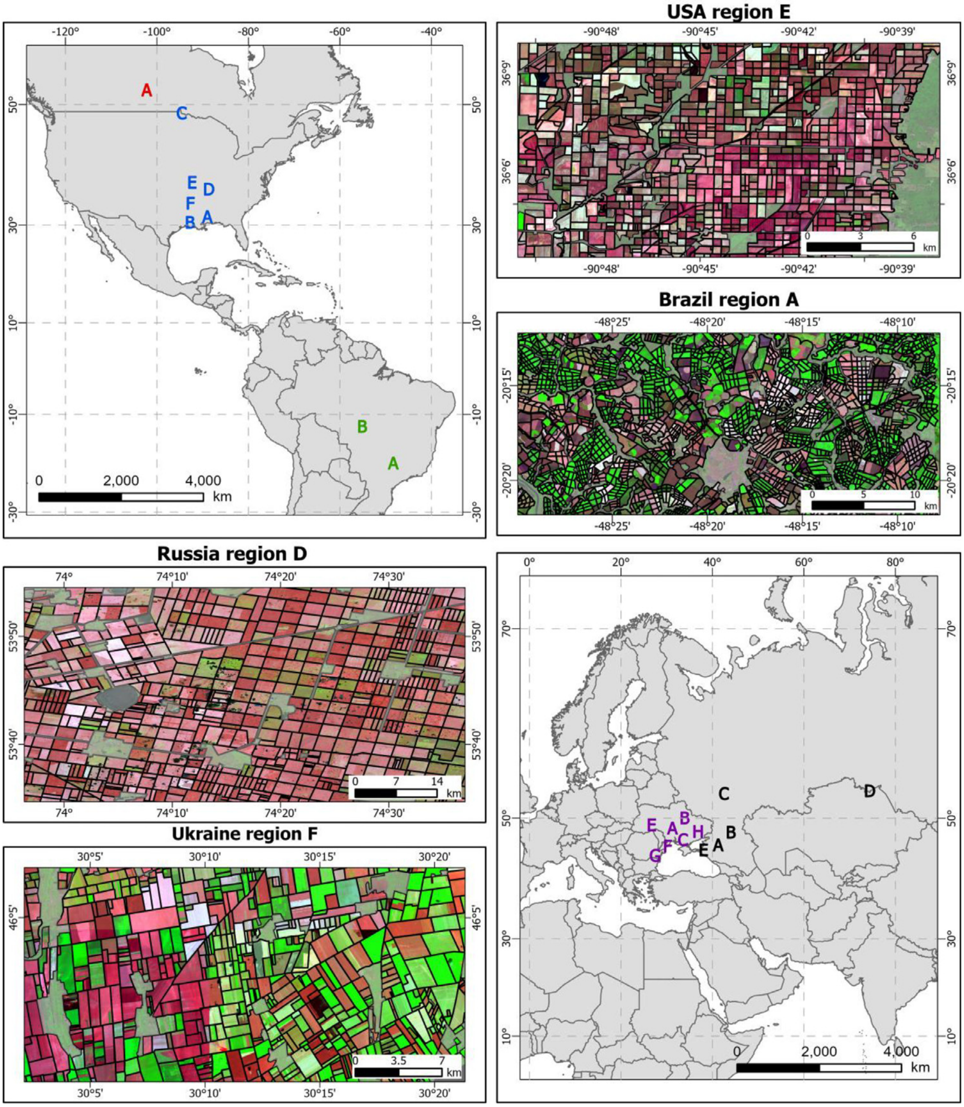

Fig. 1. Location of the twenty-two mapped regions with examples of the digitized field boundaries. For visual distinction, the colored letters represent each individual country: green (Brazil), blue (USA), red (Canada), purple (Ukraine), and black (Russia).

3. Data Description

The GloCAB cropland field boundary dataset comprises of a shapefile (and its ancillary files) containing 190,832 vectorized fields (i.e., polygons) for 22 regions spanning 5 different countries (see Fig. 1 and Table 1):

GloCAB_field_boundaries shapefile (.cpg, .dbf, .prj, .sbn, .sbx, .shp, .xml, .shx)

- Shapefile containing vectorized field boundaries for the 22 regions
- Attribute Table: Area in km^2 [*Area_km2*], Region name [Ctry_*Reg*], NoArea flag [*No_Area*], predominant crop type [*CropType*].

Table 1
Summary information on the mapped field boundary areas located in Fig. 1.

Country ID	Mapping Date	Predominant Crop Type	Cropland Field Boundaries
Brazil_A	Aug 2019	Sugarcane	4510
Brazil_B	Jul 2019	Maize	1218
Canada_A	May 2018	Spring Wheat	569
Russia_A	Jul 2019	Winter Wheat	1739
Russia_B	Aug 2019	Winter Wheat	2294
Russia_C	Apr 2019	Spring Wheat	1115
Russia_D	Apr 2019	Spring Wheat	1362
Russia_E	Oct 2018	Winter Wheat	2613
Ukraine_A	Mar 2017	Maize	3994
Ukraine_B	Mar 2017	Maize	6166
Ukraine_C	Jul 2017	Winter Wheat	9327
Ukraine_D	Aug 2016	Winter Wheat	5212
Ukraine_E	Jul 2017	Winter Wheat	5433
Ukraine_F	Jun 2017	Winter Wheat	2757
Ukraine_G	Jun 2017	Winter Wheat	10,305
Ukraine_H	Jul 2020	Winter Wheat	123,671
USA_A	Nov 2018	Sugarcane	1091
USA_B	Oct 2019	Sugarcane	2402
USA_C	Apr 2018	Spring Wheat	1342
USA_D	Sep 2020	Rice	746
USA_E	Sep 2017	Rice	1499
USA_F	Sep 2017	Rice	1467

4. Experimental Design, Materials and Methods

Each field boundary feature was manually digitized by a geospatial analyst using 20 m Sentinel-2 imagery and quality-checked by a senior team member. Generally, field boundaries were identified by hedges and similar features, which typically form semi-natural boundaries in agricultural landscapes. However, for some fields, the analysts used the productive area to delineate the boundaries, especially when fields contained distinct crop types and characteristics during the temporal window. In instances where the fields were too small to identify, the analyst digitized a larger boundary encompassing several small fields and designated the polygon with a “NoArea” flag in the attribute table. The “NoArea” designation is a flag used to remove those fields from field area statistical calculations since they do not represent the actual field boundary. In the larger GloCAB study, each region was assigned a specific temporal window (see Table 1) and all available Sentinel-2 and Planet imagery were used for that purpose. Since field boundaries do not change within a one-month timespan, the analysts digitized the perimeter using any Sentinel-2 image in that temporal window.

The predominant crop type attribute data were assigned using a combination of the GEO Global Agricultural Monitoring (GEOGLAM) Best Available Global Crop-Specific Maps (BACS) (winter wheat, spring wheat, maize, and rice) [6,7] and the 2010 Spatial Production Allocation Model (SPAM) global sugarcane physical area [8].

Limitations

In instances where fields were too small to identify individually, the analyst digitized a larger boundary encompassing several small fields and designated the polygon with a “NoArea” flag in the attribute table. Various factors impacted the analysts’ ability to accurately digitize these very small fields and observe changes over the temporal window. Examples of these factors include the time of year, soil color, crop type, and land use (e.g., small shareholder garden plots) (Fig. 2). Users should be aware that these “NoArea” polygons do not represent the true field boundaries. While it is challenging to provide a precise quantitative threshold for determining

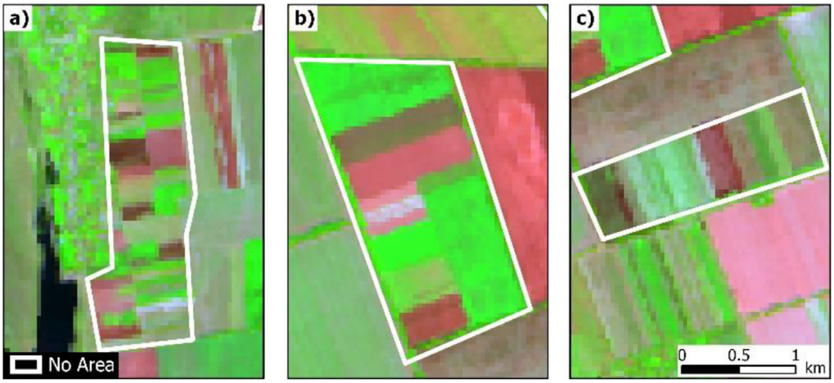

Fig. 2. Examples of "NoArea" polygons in Ukraine Region F.

when fields are too small to digitize, we have provided the size of the smallest field polygon that was mapped to offer some context: 0.0015km^2.

The predominant crop type information was extracted from two coarse resolution datasets: GEOGLAM BACS and SPAM. Given their global extent and availability, these datasets were chosen as part of the larger study [1] despite their infrequent updates and lower resolution compared to the Sentinel-2 imagery used for field delineation. We acknowledge that this may introduce some inaccuracies in identifying the predominant crop types. Therefore, we recommend that users utilize higher resolution regional or local crop type information where available to improve accuracy and reliability.

Ethics Statements

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

GloCAB Cropland Field Boundary Dataset (Original data) (Zenodo).

CRediT Author Statement

Joanne V. Hall: Conceptualization, Data curation, Supervision; **Fernanda Argueta:** Data curation, Visualization; **Louis Giglio:** Funding acquisition, Writing – review & editing.

Acknowledgments

Funding: This work was supported by NASA grants 80NSSC18K0619, 80NSSC18K0739, 80NSSC21K1961, 80NSSC21K1657, and 80NSSC18K0179, and United States Department of the Air Force contract FA8810–18-C-0017.

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.

References

[1] J.V. Hall, F. Argueta, M. Zubkova, Y. Chen, J.T. Randerson, L. Giglio, GloCAB: global cropland burned area from mid-2002 to 2020, Earth Syst. Sci. Data 16 (2) (2024) 867–885.
[2] L.D. Estes, S. Ye, L. Song, B. Luo, J.R. Eastman, Z. Meng, Q. Zhang, D. McRitchie, S.R. Debats, J. Muhando, A.H. Amukoa, High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales, Front. Artif. Intell. 4 (2022) 744863.
[3] D. Liu, F. Xia, Assessing object-based classification: advantages and limitations, Remote Sens. Lett. 1 (4) (2010) 187–194.
[4] F. Waldner, F.I. Diakogiannis, Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network, Remote Sens. Environ. 245 (2020) 111741.
[5] F. Waldner, F.I. Diakogiannis, K. Batchelor, M. Ciccotosto-Camp, E. Cooper-Williams, C. Herrmann, G. Mata, A. Toovey, Detect, consolidate, delineate: scalable mapping of field boundaries using satellite images, Remote Sens. 13 (11) (2021) 2197.
[6] I. Becker-Reshef, C. Justice, B. Barker, M. Humber, F. Rembold, R. Bonifacio, M. Zappacosta, M. Budde, T. Magadzire, C. Shitote, J. Pound, Strengthening agricultural decisions in countries at risk of food insecurity: the GEOGLAM crop monitor for early warning, Remote Sens. Environ. 237 (2020) 111553.
[7] A.K. Whitcraft, I. Becker-Reshef, C.O. Justice, L. Gifford, A. Kavvada, I. Jarvis, No pixel left behind: toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework, Remote Sens. Environ. 235 (2019) 111470.
[8] International Food Policy Research InstituteGlobal Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0, Harvard Dataverse, 2019 V4, doi:10.7910/DVN/PRFF8V.