Metadata

Dataset title: SLU historisk jordbruksmark 1942-1988/ SLU historical arable land 1942-1988

Creators: Anneli Ågren

Responsible organisation: Sveriges lantbruksuniversitet/Swedish University of Agricultural Sciences

License: CC BY 4.0

DOI: https://doi.org/10.5878/95bj-4848

Format: GeoTIFF (LZW-compression)

Bands: 1

Datatyp (Pixel type): 8-bit unsigned integer (Byte)

Lagringsdjup (Bit depth): 2 bitar per pixel (NBITS=2)

File size: 127 Mb

Coordinate system: SWEREF 99 TM (EPSG:3006)

Resolution: 10 m

Short description of dataset:

This is a dataset of historical arable land in Sweden (1942-1988). Before this study, geographical data on former arable land was only available from 2003 and onward. To improve this situation, a machine learning model was developed using Extreme Gradient Boosting (XGBoost) that could automatically detect arable land from an historical orthophoto map of Sweden (economic maps, 1:10,000 and 1:20,000). In the original maps arable land was marked with a yellow mask on top of the background orthophotos. Since the yellow mask was transparent and the background colors vary, the arable land could take on many hues. The Economic map series were downloaded from Swedish Land Survey (Lantmäteriet). (https://www.lantmateriet.se/sv/geodata/vara-produkter/produktlista/ekonomiska-kartan/). A machine learning model was trained to identify arable land. Overall, the XGBoost model demonstrated robust performance. The map then underwent post-processing steps and was aggregated to 10 × 10 m. The dataset is a mosaiced raster with a spatial resolution of 10 × 10 m, representing arable land extracted from 12,015 and 1,046 economic maps at scales of 1:10,000 and 1:20,000, respectively. It shows arable land in the maps that were produced 1942-1988. Note that the arable land delineation is made from orthophotos taken a few years earlier. The post processing steps affected the quality of the map.

The overall quality of the final 10 × 10 m 'SLU historical arable land map 1942-1988' was assessed across a transect from north to south. The quality measures were all high with Cohen's Kappa and Matthews Correlation Coefficient of 0.974. For the not arable class, precision was 0.977, recall was 0.997, and F1-score was 0.987. The arable class achieved a precision of 0.997, a recall of 0.977 and F1-score of 0.987.

The map covers 404,975 km^2, of which 32,535 km^2 is historical arable land. Each cell is assigned a value of one for arable land and zero for all other land uses and surface waters. Areas where no map could be produced was assigned NoData (the NoData value is set to 2). In areas covered by more than one Economic map, the oldest version was selected to extend the temporal range of the dataset as far back as possible (see the Economic Maptile Overview shapefile for individual map production years).

The dataset is released under a CC BY 4.0 license, allowing unrestricted use, provided that it is cited.

To cite datset use: Ågren, A. M. SLU historisk jordbruksmark (Version 1) [Data set]. Sveriges lantbruksuniversitet. https://doi.org/10.5878/95bj-4848 (2026).

In depth data description:

A full method development article including the statistical evaluation of the map quality is published as a Data Descriptor in the journal Scientific Data.

Dataset organisation:

The complete GeoTIFF raster dataset comprises several files with the same name but different file extensions.

.tif — The main raster image (127 MB). Contains the actual pixel data showing historical arable land

Performance & display

- .tif.ovr (Overview/Pyramid file) — Pre-rendered lower-resolution versions of the image. This is what lets the map zoom out smoothly without reading all 127 MB every time. Without it, zooming out is very slow.

Metadata

- .tif.aux.xml — Stores statistics (min/max values, histograms, color interpretation) calculated by software like ArcGIS or GDAL. Speeds up rendering and styling.
- .tif.aux.xml.xml — Another XML file with raster statistics and a histogram.
- .tif.ovr.aux.xml — Statistics for the overview raster.

Attribute table

- .tif.vat.dbf — A Value Attribute Table linking each pixel value 0=NotArable and 1=Arable. Makes the data meaningful.
- .tif.vat.cpg — Specifies the text encoding for the .dbf file so characters display correctly.
- .clr – A colormap for ArcGIS with suggested arable land in yellow and green for non-arable land. Change colors as you please.

In short: the .tif is the data, the .ovr makes it fast, and the rest describe what the values mean.

The data files can be read using a GIS-program such as ArcGIS, QGIS or GDAL.

QGIS

- Open directly via *Layer → Add Layer → Add Raster Layer* or just drag and drop the .tif into the canvas
- QGIS will automatically pick up the .ovr, and .aux.xml files in the background
- Free and open source — handles GeoTIFFs natively via GDAL

ArcGIS

- Open in ArcMap or ArcGIS Pro via *Add Data*, or drag into the map view
- The .vat.dbf attribute table is especially useful here — ArcGIS will display it in the layer's attribute table automatically
- The .ovr pyramids are also natively supported

A few practical tips:

- Always keep all files in the same folder — if you move just the .tif, the software may lose the georeferencing or slow down significantly.
- At 127 MB, the .ovr pyramid file matters a lot — if it's missing, programs will ask if you want to build pyramids before rendering efficiently.
- The .vat.dbf and .clr are more ArcGIS-centric. In ArcGIS you can use the preset color legend .clr which colors the arable fields yellow and all else transparent, but feel free to choose you own colors. In QGIS you need to set up a color map or legend manually based on the pixel values.

EconomicMaptileOverview

EconomicMaptileOverview is a polygon shapefile showing the footprints and filenames of the generalized economic map sheets considered for the historical arable land dataset. It contains 13,061 polygons: 12,015 sheets at a scale of 1:10,000 and 1,046 at a scale of 1:20,000. Each polygon represents the complete extent of a map-sheet raster, not the extent of mapped arable land. Some sheet footprints overlap due to imperfect georeferencing (mostly 1:20 000). Map production years vary across the study area. According to the dataset's mosaicking procedure, the oldest available map was selected for each location.

For 24 filenames without an identifiable year, Year is set to 0 and YearStatus to Missing; zero is a missing-value marker, not a production year.

Attribute table for EconomicMaptileOverview

FID	The ID number of the polygon
Filename	Full filename of the corresponding map-sheet TIFF.
Year	Production year extracted from the filename; 0 indicates a missing year.

YearStatus “Extracted” if a year was found otherwise “Missing”.

MapScale Map scale: 1:10,000 or 1:20,000.

Coordinate system and files

The overview is provided in SWEREF 99 TM (EPSG:3006).

The ESRI shapefile consists of several files that should be kept together:

File extension:	**Description:**
.shp	Polygon geometry of the map-sheet footprints.
.shx	Index linking records to the geometries.
.dbf	Attribute table.
.prj	Coordinate reference system.
.cpg	Character encoding for the attribute table.
.sbn and .sbx	Optional spatial-index files that improve GIS display and query performance.
.xml	Optional ArcGIS metadata file.

Code Availability:

The code for the XGBmodel and the production of the map can be found on GitHub.
https://github.com/anneliagren/HistoricalArableLandSweden

Data overview:

An overview of the dataset is shown in Figure 1 below.

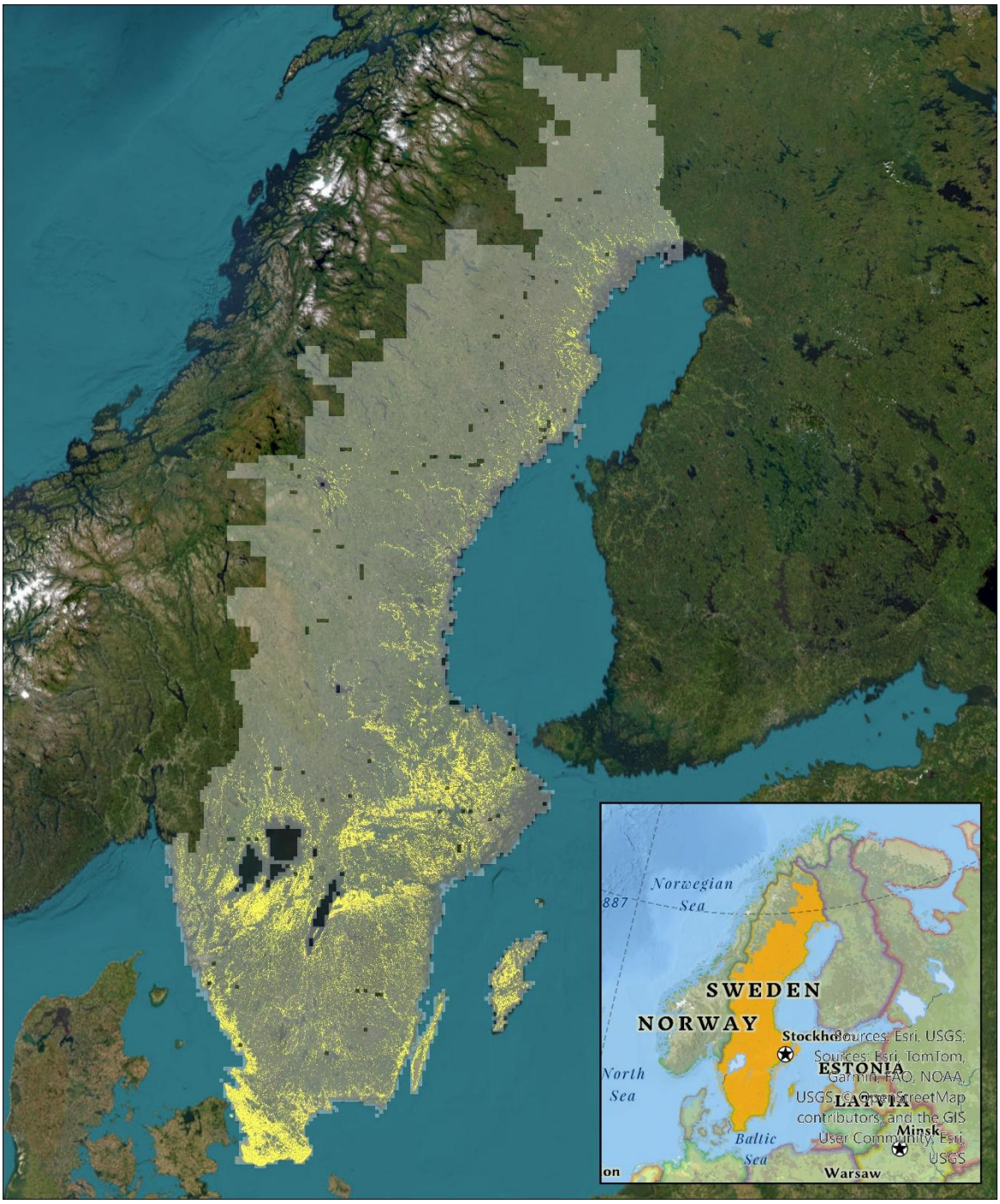


Figure 1. Overview of the study area, light colored area overlaying satellite image and orange in inset map. The historical arable land from the historical economic maps is denoted in yellow. The "SLU historical arable land" map covers most of Sweden, with the exception of the mountain range in northwestern Sweden, where no economic maps were produced. Some additional areas are missing because the original map files were unavailable or corrupted, or because they were manually excluded due to unreasonable arable land classifications produced by the XGBoost model. All areas for which no map could be produced were assigned NoData values. Satellite data is from Esri, Maxar, Earthstar Geographics, and the GIS User Community is CC0 open data and the final dataset is released under a CC BY 4.0 attribution license, allowing unrestricted use, provided that it is cited.