Yield data – arable / cereals

Guidance on using agricultural datasets in archaeological and heritage management applications

Technical and conceptual notes to support data exchange via API platforms and Ecosystem Services Platforms (online GIS platforms)

Yield data quantifies the amount of crop harvested in different areas within a field. This data is collected using monitors attached to combine harvesters with an attached GPS/GNSS system. These measurements are typically acquired real time and stored for later analysis and can be mapped spatially. There are several methods for yield measurement which provide different quantifications, but all of them provide information on spatial variability in the amount of crop harvested within a field. Yield is an output of everything that happened during the cropping season, including soil type, weather conditions, management practices, disease, pests, and interactions between all of these factors, and the outcomes of these interactions are also related to the appearance of archaeological cropmarks. Yield monitoring is most often applied to arable, cereals crops. It may be collected repeatedly across multiple years to assess the impacts of management decisions or changing environmental factors.

Data Type

Vector point measurements

Text file or vector file - 'Raw' or 'cleaned' yield measurements are reported as an irregularly spaced collection of points with associated attributes. Different instruments will vary in the format of the data provided and will measure different variables, depending on their configuration. Typical file types include .txt, .csv, .json, .geojson, and shapefiles.

Common Data Derivatives

Yield maps – cleaned and interpolated georeferenced raster data, with calibrated yield values represented.

Management zones – polygonal areas delimited as significantly different in yield from neighboring areas in the field

Identifiers

Yield data representing in field variability is not widely available in public repositories. Individual farmers, cooperatives, advisory services, and land managers must be approached.

Vocabulary (thematic tags)

(Thematic tags, term lists, thesauri)

The following tags are recommended to provide compatibility across archaeological and agricultural vocabularies.

AGROVOC: plant response http://aims.fao.org/aos/agrovoc/c_25446; crop monitoring http://aims.fao.org/aos/agrovoc/c_37838

GEMET: multiple use management area https://www.eionet.europa.eu/gemet/en/concept/5424

Getty AAT: Crop marks http://vocab.getty.edu/page/aat/300248611; landscape archaeology http://vocab.getty.edu/page/aat/300252459; Landscapes (environments) http://vocab.getty.edu/page/aat/300008626

For UK collections: Evidence (England) CROPMARK ; Event Type (England) EVALUATION

Data structure – single harvest yield data

For any point datasets representing in-field variability in yield

Data structures for yield data vary by manufacturer. The information contained in a file, field names, and field order will vary. The field names may or be integrated into the data file, provided in a separate header file, or known by consulting the system's documentation.

An example of a 'raw' data structures is:

"Full": ddd.dddddd,dd.dddddd,mm.mm,tttttttt,n,lll,www,cc.c, kk,ppppp,ssssss,Fnn:bbbbbbbb,Lnn:bbbbbbbb,gggggggggg, sss, ppp,aaaa

where: ddd.dddddd = longitude (degrees, + East and - West) dd.dddddd = latitude (degrees, + North and - South) mm.mm = grain mass flow (pounds per second) tttttttt = GPS time (seconds) n = cycle period (seconds) lll = distance travelled in cycle period (inches) www = effective swath width (inches) cc.c = moisture content (percent wet basis) kk = status (bits 0 thru 4 - header down) ppppp = pass number ssssss = yield monitor serial number Fnn:bbbbbbbb = field ID (number and name) Lnn.bbbbbbbb = load ID (number and name) gggggggggg = grain type sss = GPS status ppp = point dilution of precision (PDOP) aaaa = altitude (feet) A typical exported data string might look as follows, -85.238446,38.308450,8.36,838257850,1,54,348,28.3, 33,0, 950304,"F4:901","L1:CLARK","Wheat",12,64,813

(example from https://lter.kbs.msu.edu/datasets/40, John Deere combine format)

Data in column order Description 1. Longitude Decimal degrees 2. Latitude Decimal degrees 3. Flow Pounds per second 4. GPS Time Seconds 5. Logged Interval Seconds 6. Distance Inches 7. Swath Inches 8. Moisture Percent (wet basis) 9. Head Status 1 = harvesting, 0 = not harvesting 10. Pass Number Generally +1 for each header up/down 11. Any other information Note: The file must have 11 columns at a minimum. The content of columns 11 and higher will not affect the reading and processing of the data but the 11th column is needed.

(example of required column sequence for Yield Editor, Ag Leader Advanced Format).

These data may be processed before delivery to the end user. An example processed data structure is:

"Basic": ddd.dddddd,dd.dddddd,yyy.y,cc.c,ssssss,Fnn:bbbbbbbb, Lnn:bbbbbbbb,gggggggggg

where: ddd.dddddd = longitude (degrees, + East and - West) dd.dddddd = latitude (degrees, + North and - South) yyy.y = grain yield (marketable bushels per acre) cc.c = moisture content (percent wet basis) ssssss = yield monitor serial number Fnn:bbbbbbbb = field ID (number and name) Lnn.bbbbbbbb = load ID (number and name) gggggggggg = grain type.

(example from https://lter.kbs.msu.edu/datasets/40)

A variety of extensions for yield data files are used, which are normally text files following a specific format.

Table 1. File Formats Generated by Various Yield Monitor Displays.

From: https://ohioline.osu.edu/factsheet/anr-8

Note that many file formats are proprietary and a software license may be required to translate them into an open geospatial format. Common data formats are CLAAS, John Deere and New Holland.

Company	**In-cab Display / System**	**File Format**
Ag Leader	YM2000, PF3000 and PFAdvantage	*.yld
	Insight	*.ilf
	Integra, InCommand and newer	*.agdata (individual files represent a field)
AGCO	FieldStar II	TaskData.xml
Case AFS	AFS Pro600 and Pro700	*.yld;*.vy1 Hierarchy of file directories, typically beginning with an upper level file directory named for the date it was created… 141119M7.cn1\...
CLAAS	CEBIS Quantimeter	.dat
John Deere	GreenStar 1 GreenStar 2 and newer	*.gsd; *.gsy *.ver Hierarchy of file directories, typically beginning with an upper level file directory like… GS3_2630\harvest\RCD…
New Holland	Intelliview	*.vyg, *.vy1, *.yld
Precision Planting	20/20	*.dat Data can be synchronized continuously to the "cloud" through Climate FieldView.
Topcon	YieldTrakk	ISOXML

Core attributes of data types (see above)

Critical information for inclusion

- Raw point data – point values including gps locations and attributes including dry yield.

- Cleaned point data – a geolocated or georeferenced point file where each point's attributes contain the numeric values with calibration and cleaning applied; extent is defined by a bounding box in a WGS84 degrees or UTM metres
- Processed raster data – yield maps which represent continuous yield values, based on the interpolation of cleaned point datasets; extent is defined by a bounding box in a WGS84 degrees or UTM metres
- Vector Interpretation – Polygonal areas created to define management zones which delimit areas of the field which can be managed uniformly based on a range of factors.
- (Where available) Calibration data – Sample values including attributes; may not be geospatial.

Data Processing

Raw yield data recommended processing - calibration:

- Yield monitors should be calibrated to produce accurate yield data and enable year-on-year comparisons. These calibrations can take place during data collection. There are several types of calibration data, comprising: mass-flow sensor vibration calibration, temperature calibration, moisture sensor calibration, mass flow sensor (weight) calibration, and dGPS calibration. Calibrations are carried out per crop type. There may be multiple calibrations carried out during a single harvest, with the number dependent on the degree and frequency of changes in weather conditions or overall crop moisture sensed. In practice, no calibration, improper or inconsistent calibration is common, due to the time required and degree of complexity of the process.
- If calibration data is supplied separate to the yield measurements, calibrations should be applied following equipment manufacturer's instructions.

Raw yield data recommended processing - basic cleaning:

- Correct for GPS/GNSS signal delay / timing errors
- Remove points where the header was not engaged or where the combine was turning
- Correct for effects of the combine position partially overlapping previously harvested areas
- Remove points which represent measurements that are statistical outliers relative to the yield distribution at the field level.

Further cleaning suggestions are made at: https://www.aspexit.com/filtering-cleaning-yield-maps/.

Cleaned yield data recommended processing:

- Interpolation to create a raster and facilitate visual interpretation of variations in yield values.

Processed raster data:

- Rescaling of measured yield values to visually emphasize variability within the dataset.
- Creation of raster derivatives e.g. variation between years, Deviation from local mean yield Deviation from local mean yield in multiple years

Management zones:

- No further processing needed after management zones have been created.

Dataset Level Core Metadata

Point Metadata		
(All data types)		
Duration/Collection date	Mandatory	Date (ISO 8601)
Instrumentation details	Mandatory	Character String (Free text)
Field Name	Mandatory	Character String (Free text)
Bounding Box	Mandatory	Coordinates (WGS84, decimal degrees)
Nominal Spatial Resolution	Mandatory	Decimal / Numerical
Weather data reference	Mandatory	Character String (Free text)
Crop type	Mandatory	Character Strong (controlled vocabulary)
Calibration	Optional	Character String (Free text)
Processing	Optional	Character String (Free text)

Raster Metadata		
(All data types)		
Duration/Collection date	Mandatory	Date (ISO 8601)
Instrumentation details	Mandatory	Character String (Free text)
Field Name	Mandatory	Character String (Free text)
Bounding Box	Mandatory	Coordinates (WGS84, decimal degrees)
Nominal Spatial Resolution	Mandatory	Decimal / Numerical
Interpolation type	Mandatory	Character String (Free text)
Crop type	Mandatory	Character Strong (controlled vocabulary)
Related data reference	Mandatory	Character String (Free text)

Management Zones Metadata		
(All data types)		
Field Name	Mandatory	Character String (Free text)
Bounding Box	Mandatory	Coordinates (WGS84, decimal degrees)
Crop type	Mandatory	Character Strong (controlled vocabulary)
Related data reference	Mandatory	Character String (Free text)

Note that diverse standards for yield data have been proposed and are actively being developed. As this is a dynamic space, the metadata recommendation is likely to change. See https://www.aspexit.com/standards-and-data-exchange-in-agriculture/ for an overview and https://github.com/opengeospatial/Agriculture-DWG as an example initiative.

Scope Note

Applications of precision agricultural yield measurements in archaeology

Yield data shows in-field variations in the final yield of a crop and may contain additional information on in-field variability in moisture contained in the crop. These data are typically at a relatively coarse spatial resolution with a circa 10m sampling interval. Yield data is unlikely to reveal individual small archaeological features. However, yield data may reveal larger individual archaeological features or indicate areas with a concentration of smaller features which have an aggregate effect on crop development. In crops which do not produce visible cropmarks observable from aerial platforms, yield data across multiple years (analysis of historical series) may serve as a proxy for cropmarks: areas of persistent variability in yield across a field despite treatments may be indicative of the potential presence of buried archaeological features. Yield data may be combined with observations of cropmarks or locations of known archaeological features to assess the interaction between crop root systems and archaeological deposits. Yield data should be interpreted with reference to information on agricultural interventions (e.g. irrigation and fertilisation), weather records across multiple years, and available soil survey data.

Scope Note

Archaeological influences on measured yield to consider in Precision Agriculture

Where crop root systems interact with archaeological deposits in the shallow subsoil or topsoil, the presence of buried archaeology may explain some in-field variations in yield. Where cropmarks have been observed in past years or the present year, yield is more likely to be impacted by buried archaeological deposits. In these situations, models of predicted yield and assessments of yield could be improved by accounting for the likelihood of increased local variability in SMD, shallow subsoil structure, and micronutrients.

Relevant Literature

Corti, M., Marino Gallina, P., Cavalli, D., Ortuani, B., Cabassi, G., Cola, G., Vigoni, A., Degano, L., & Bregaglio, S. (2020). Evaluation of In-Season Management Zones from High-Resolution Soil and Plant Sensors. *Agronomy*, *10*, 1124. https://doi.org/10.3390/agronomy10081124

Hejcman, M., & Smrž, Z. (2010). Cropmarks in stands of cereals, legumes and winter rape indicate sub-soil archaeological features in the agricultural landscape of Central Europe. *Agriculture, Ecosystems & Environment*, *138*(3), 348–354. https://doi.org/10.1016/j.agee.2010.06.004

Hejcman, M., Součková, K., & Gojda, M. (2013). Prehistoric settlement activities changed soil pH, nutrient availability, and growth of contemporary crops in Central Europe. *Plant and Soil*, *369*(1), 131–140. https://doi.org/10.1007/s11104-012-1559-y

Kc, K., Zhao, K., Romanko, M., & Khanal, S. (2021). Assessment of the Spatial and Temporal Patterns of Cover Crops Using Remote Sensing. *Remote Sensing*, *13*(14), 2689. https://doi.org/10.3390/rs13142689

Khanal, S., Fulton, J., & Shearer, S. (2017). An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture*, *139*, 22–32. https://doi.org/10.1016/j.compag.2017.05.001

Leroux, C., Jones, H., Taylor, J., Clenet, A., & Tisseyre, B. (2018). A zone-based approach for processing and interpreting variability in multi-temporal yield data sets. *Computers and Electronics in Agriculture*, *148*, 299–308. https://doi.org/10.1016/j.compag.2018.03.029

Leroux, C. (2021) *Quantifying the within-field heterogeneity or variability in agriculture – Aspexit*. Retrieved 27 September 2021, from https://www.aspexit.com/quantifying-the-within-field-heterogeneity-or-variability-in-agriculture/

Loisel, P., Charnomordic, B., Jones, H., & Tisseyre, B. (2019). An optimisation-based approach to generate interpretable within-field zones. *Precision Agriculture*, *20*(1), 101–117. https://doi.org/10.1007/s11119-018-9584-3

Taylor, J. A., Tisseyre, B., & Leroux, C. (2019). A simple index to determine if within-field spatial production variation exhibits potential management effects: Application in vineyards using yield monitor data. *Precision Agriculture*, *20*(5), 880–895. https://doi.org/10.1007/s11119-018-9620-3