

## EODal: An open-source Python package for large-scale agroecological research using Earth Observation and gridded environmental data

Lukas Valentin Graf<sup>a,b,\*</sup>, Gregor Perich<sup>a</sup>, Helge Aasen<sup>a,b</sup>

<sup>a</sup> Crop Science, Institute for Agricultural Science, ETH Zurich, Universitätsstrasse 2, CH-8092 Zürich, Switzerland

<sup>b</sup> Earth Observation of Agroecosystems Team, Devision Agroecology and Environment, Agroscope, Reckenholzstrasse 191, CH-8042 Zürich, Switzerland

### ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.7278252>

#### Keywords:

Satellite data

Python

Open-source

Earth Observation

Ecophysiology

### ABSTRACT

Earth Observation by means of remote sensing imagery and gridded environmental data opens tremendous opportunities for systematic capture, quantification and interpretation of plant–environment interactions through space and time. The acquisition, maintenance and processing of these data sources, however, requires a unified software framework for efficient and scalable integrated spatio-temporal analysis taking away the burden of data and file handling from the user. Existing software products either cover only parts of these requirements, exhibit a high degree of complexity, or are closed-source, which limits reproducibility of research. With the open-source Python library EODal (Earth Observation Data Analysis Library) we propose a novel software that enables the development of fully reproducible spatial data science chains through the strict use of open-source developments. Thanks to its modular design, EODal enables advanced data warehousing especially for remote sensing data, sophisticated spatio-temporal analysis and intersection of different data sources, as well as nearly unlimited expandability through application programming interfaces (APIs).

### 1. Introduction

Images from Earth Observation (EO) satellites and in-situ observations are of great importance for ecophysiological (Caparros-Santiago et al., 2021) and agroecological research (Karthikeyan et al., 2020). Such data can be used to determine plant traits and allow mapping of plant growing conditions for larger areas using standardized methods (Weiss et al., 2020). Open-access, high-resolution satellite data such as from the European Space Agencies' Sentinel-2 (S2) mission can resolve field heterogeneity and provide site-specific farming measures operationally. Examples include yield estimates (Marshall et al., 2018; Perich et al., 2022), extraction of phenological metrics (Duarte et al., 2018), variable irrigation rates (Barker et al., 2018) and site-specific fertilization scheduling (Mittermayer et al., 2022). Remotely sensed plant traits and their dynamic development over time can further be augmented with environmental covariates such as climate, soil and terrain data, as well as information about farm management to perform integrated analysis on material and energy fluxes across spatio-temporal scales (Asam et al., 2018).

However, accessing, managing and analyzing EO data is complex and often requires solid knowledge of geographic information science and coding to properly handle large spatial data sets. We base

this finding on an exhaustive review of existing software tools. For example, the philosophy of the OpenDateCube<sup>1</sup> initiative is about integrated analysis of EO data using standardized interfaces. Installing the software, however, is complex as setting up a database instance is required. The Framework for Operational Radiometric Correction for Environmental monitoring (FORCE) (Frantz, 2019) is primarily designed for the creation of Analysis-Ready-Data (ARD) but does not provide interfaces for data analysis workflows. Analysis workflows using standardized interfaces are the main subject of the openEO<sup>2</sup> project. openEO, however, focuses on cloud environments. Thus, data sets that are not available on externally operated web platforms are currently excluded. This applies particularly to (experimental) research data sets. Researchers therefore often spend a significant amount of time getting the data into an analysis-ready format. Based on the analysis of more than 3000 productive machine learning pipelines at Google, Xin et al. (2021) identified great potential for optimization in the area of data management and pre-processing, which also appears to be true in the EO area.

For these reasons, we developed the Earth Observation Data Analysis Library (EODal) as an open-source Python (3.8+) package designed to make EO data analysis tools available to researchers without the

\* Corresponding author at: Crop Science, Institute for Agricultural Science, ETH Zurich, Universitätsstrasse 2, CH-8092 Zürich, Switzerland.  
E-mail address: [lukasvalentin.graf@usys.ethz.ch](mailto:lukasvalentin.graf@usys.ethz.ch) (L.V. Graf).

<sup>1</sup> <https://www.opendatacube.org/>

<sup>2</sup> <https://openeo.org/>

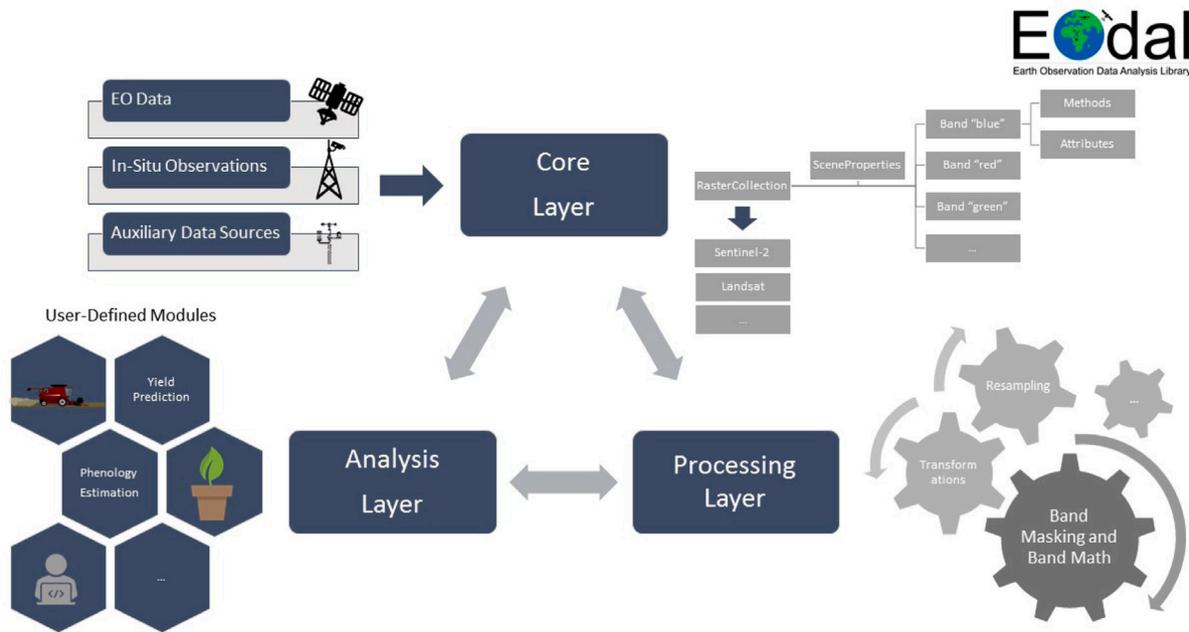


Fig. 1. Overview of EOdal and its three layers: The core layer (top) handles different input data sources in a standardized way. The analysis layer (bottom left) builds upon the core and the processing layer (bottom right) for data maintenance, querying, intersection and integration into user-defined data science pipelines.

need for in-depth knowledge about geoinformation science, coding and remote sensing data handling. A key aspect of EOdal is the ability to apply spatial data science methods to data from different sources within a unified framework based on open-source tools. We also intend to provide researchers with an alternative to proprietary software solutions such as the widely used Google Earth Engine (Gorelick et al., 2017).

The functionality and structure of EOdal is explained in Section 2 of this paper. In Section 3, we present a reproducible use case based on an agricultural research question followed by a discussion in Section 4.

## 2. Description of EOdal

EOdal consists of three layers as shown in Fig. 1. The layers are organized in a triangle to emphasize their inter-dependencies. The core layer (Fig. 1, top) provides the Python classes required to perform I/O operations to read and write geo-spatial data sets in a generic way. It is also the basis for data warehousing, i.e., the storage of metadata. Class inheritance extends its capabilities to specific EO sensors such as S2 Multispectral Imager (e.g., for convenient reading of data organized in the Satellite Archive for Europe structure). (Pre-)processing of these datasets is accomplished in the processing layer (Fig. 1, lower right). Processing steps such as spatial reprojection are often a necessity to combine different datasets for analysis. The analysis layer (Fig. 1, lower left) enables automatized, reproducible EO data management and is the backbone for data-driven analysis of geo-spatial data sets and their spatio-temporal intersection.

Deployment of EOdal is independent of an Operating System. Furthermore, EOdal can be used on local premises (e.g. for processing research datasets) but also in cloud environments for fast access to large, freely accessible data such as the global S2 or Landsat archive. To enable fast and scalable deployment required for large-scale analysis tasks EOdal can be installed into containerized environments (Docker containers).

### 2.1. Core layer

The data model of the core layer (Fig. 1 top) follows the object-oriented programming paradigm and includes different classes. There

are three classes in the EOdal core layer: Bands, RasterCollections and derived (inherited) sensor-specific classes. The Band class represents the base class. In simple terms, a Band refers to a two-dimensional array referenced in a geo-spatial coordinate reference system. A RasterCollection is a collection of zero to  $n$  Band objects, which can exist in different spatial reference systems, grid cell sizes (pixel sizes) and spatial extents. Band objects in a collection are identified by names (e.g., “blue”, Fig. 1 top right) instead of numeric indices. Bands and RasterCollections can be created from any geo-referenced raster data set understood by Geospatial Data Abstraction Library (GDAL, e.g., GeoTiff), vector features (e.g. Shapefile, GeoJSON), and from numerical arrays (Python libraries, e.g. NumPy, Zarr). Sensor-specific RasterCollections include all the functionalities and attributes of the RasterCollection class, but are tailored to the requirements and capabilities of specific imaging sensors such as, for example, S2, or Landsat. The inheritance-based software design allows the introduction of further sensors and makes the core layer extensible also with regard to upcoming future EO platforms such as the hyperspectral Copernicus expansion mission CHIME<sup>3</sup> and beyond.

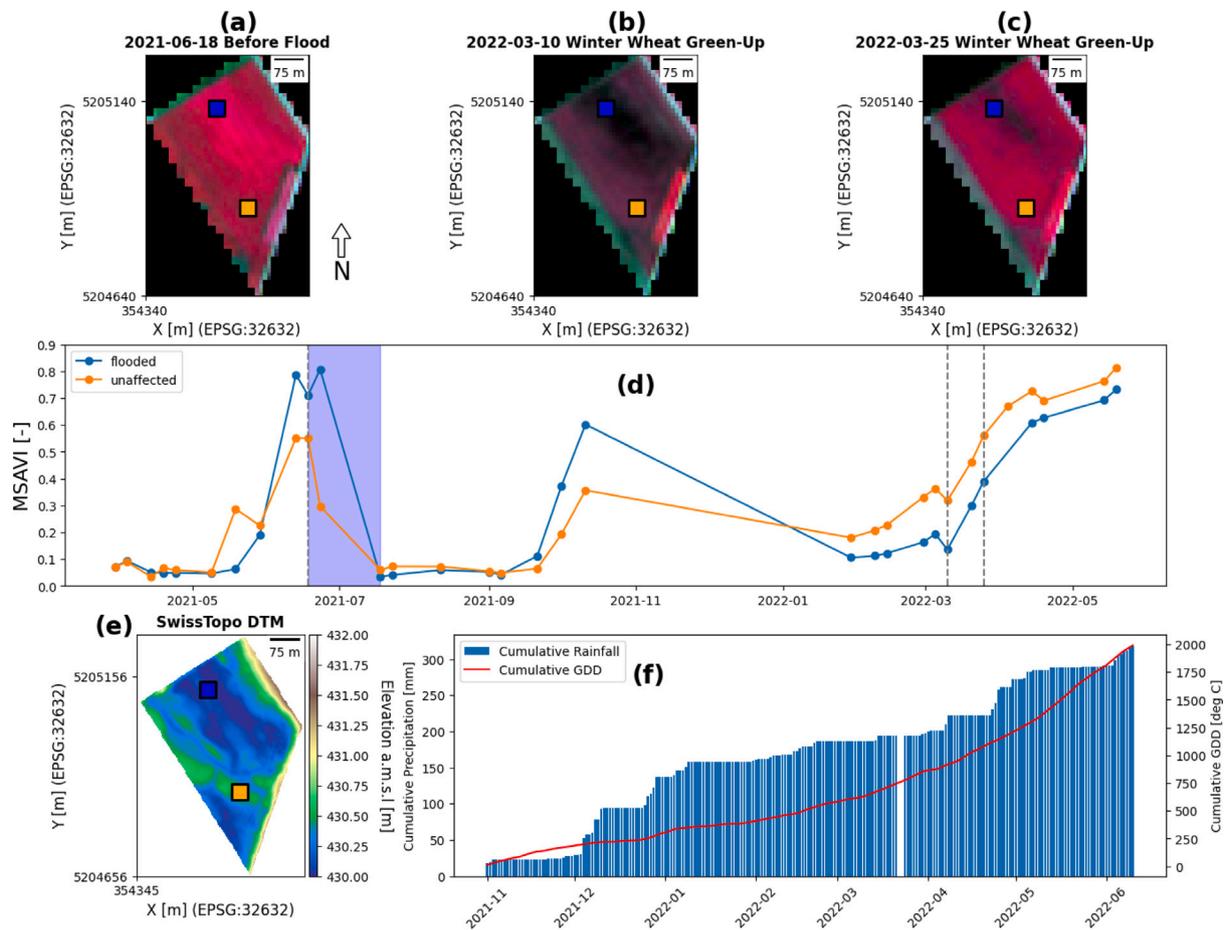
A further central element of the core layer is the collection of metadata in a spatio-temporal catalog allowing the filtering of records by data source, time period, and geographic region of interest (ROI). EOdal supports Spatio-Temporal Asset Catalogs (STAC), which are available in many cloud environments that provide geo-spatial (satellite) data, such as Microsoft Azures Planetary Computer,<sup>4</sup> Amazon Web Services (AWS) Earth<sup>5</sup> and the Copernicus Data and Information Access Services (DIAS). For local deployment, EOdal offers the possibility to store metadata in a PostgreSQL database with the spatial PostGIS extension.

In terms of spatial data models and standards, EOdal supports area (polygons and multi-polygons) as well as point features following the general feature model defined in the ISO 19109 standard (ISO, 2015). This allows point-based in-situ observations, for example from weather stations or ground sensors, to be intersected with EO data and auxiliary data sources such as Digital Elevation Models (see Section 3).

<sup>3</sup> [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Going\\_hyperspectral\\_for\\_CHIME](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Going_hyperspectral_for_CHIME)

<sup>4</sup> <https://planetarycomputer.microsoft.com/>

<sup>5</sup> <https://aws.amazon.com/de/earth/>



**Fig. 2.** Result of a single Jupyter notebook run using EOdal (v0.0.1) on Microsoft Planetary Computer. S2 false-color infra-red composite of the field parcel before the heavy rainfall events in June and July 2021 (a) and spatial heterogeneity in winter wheat green-up during spring 2022 (b–c). In addition, cumulative growing degree days (GDD) and daily precipitation sums derived from a nearby weather station are shown (f). The extent of the flood in 2021 – evident as darker areas in (b–c) – corresponds to a depression visible in the Digital Elevation Model (e). The flood affected plant growing conditions in spring 2022 as shown in the MSAVI time series (d) of a ‘flooded’ pixel (blue) and a pixel that was little affected by the floods in 2021 (orange). The dashed lines in (d) correspond to the timing of (a), (b) and (c), respectively, whereas the blue rectangle in (d) indicates the approximate timing of the flooding in 2021.

## 2.2. Processing layer

The processing layer (Fig. 1, lower right) provides functionalities for EO data preparation and (pre-)processing. Data preparation is usually necessary to enable the merging datasets from different sensors and platforms. This includes image manipulation methods such as spatial resampling and reprojection from one coordinate system into another or masking operations to mask out bad quality observations or land cover classes.

## 2.3. Analysis layer

The analysis layer (Fig. 1, bottom) is used for the analysis of data from EO satellites and environmental covariates such as meteorological, terrain, soil and land use data. It builds upon the core layer to query and access different data sources and the processing layer to prepare the data analysis-ready. The underlying complex data management such as the merging of data from satellite tiles and filling of no-data values is hidden from the user in the processing layer (see Section 2.2).

EOdal provides interfaces to widely-used open-source Python libraries (e.g., geopandas, numpy, xarray). This allows users to integrate their own EO processing workflows and modules (c.f. Section 2.4) to e.g. estimate biochemical, structural or integrated traits such as chlorophyll, leaf area index, yield or land surface phenology (Fig. 1, lower left).

## 2.4. Applications of EOdal

EOdal has already been used for studies in agroecological research. Perich et al. (2022) used EOdal for pixel-based prediction of crop yield using S2 time series. In Graf et al. (2022), EOdal is used to propagate radiometric uncertainty from S2 reflectance factors to phenological metrics. Moreover, EOdal drives the EO platform at the Swiss Federal Center of Excellence for Agricultural Research, Agroscope, underpinning its relevance within an operational and governmental environment.<sup>6</sup>

## 3. Usage example: Interpreting in-field growth heterogeneity through time with topographic and meteorological data

We illustrate here the potential of EOdal for integrated spatio-temporal analysis over an example ROI from western Switzerland near Lake Neuchâtel (46.98°N, 7.07°E). All results were produced by a single Jupyter notebook (<https://doi.org/10.5281/zenodo.7278252>) running EOdal v0.0.1 on Microsoft Planetary Computer. Due to large-scale land subsidence, the entire region – a former peat land – is subject to increased flood risk (Egli et al., 2020). An agricultural parcel (12 ha) was affected by flooding during heavy rainfall (360 mm within 30 days) between mid-June and July 2021, causing flood damage to

<sup>6</sup> <http://www.eoa-team.net/>

the previously homogeneous green canopy. Fig. 2 shows S2 derived false-color infra-red images of the parcel in 2021 before the flood (reddish tones in Fig. 2a). In the spring 2022, differences in canopy greenness of the emerging winter wheat are evident (Fig. 2b and c). Darker areas, which indicate lower soil cover, can be discerned in the otherwise red-colored canopy. The spatial pattern within the field corresponds to the patterns within the DEM (Fig. 2e) and reveals that parts with lower elevation values were more impacted by flooding than the rest of the field. Extraction of the Modified Soil Adjusted Vegetation Index (MSAVI, Qi et al., 1994) time series from two selected S2 pixels (Fig. 2d) allows to evaluate the impact of the flood in 2021 on the growth dynamic in 2022 and confirms the findings from the false-color S2 images.

As additional usage example, the growth pattern of a field is investigated in relation to meteorological data, often used to validate and calibrate ecophysiological growth models. In this case (Fig. 2f), the cumulative growing degree days (in °C) and cumulative daily rainfall amount (in mm) were derived from a nearby weather station (Ins) operated by the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) calculated from November 1<sup>st</sup> 2021.

#### 4. Discussion and conclusion

The usage example shown in Fig. 2 highlights how EOdal can be used to combine different environmental covariates with vegetation dynamics obtained from satellite time series to develop a holistic understanding of plant–environment interactions. This simple example shows how EOdal lowers the barriers of entry to EO analysis unlocking potential for data-based decision-making in agroecology. In addition, by using Docker and cloud infrastructure, the analysis can be extended to larger spatial and temporal scales. For the future we envision an ecosystem of user modules that build on EOdal, complement themselves and are developed and made freely available to the EO community.

EOdal is in line with recent developments in agriculture that allow (technically inexperienced) users to run complex analyses: Godara et al. (2022) developed a platform called “AgriMine”. The platform enables spatio-temporal analysis of issues in Indian agriculture based on help-desk calls improving the Indian agricultural extension service. Improvements in variety testing were enabled by an online platform in China that makes processes more efficient and lines up with recent incentives in high-throughput phenotyping efforts (Pan et al., 2022). Similarly, EOdal can contribute to increased use of EO data in agriculture.

Although we focus on agriculture (Section 2.4), EOdal is not limited to it. Rather, EOdal can be used in all disciplines that process EO data and require open, reproducible data analysis workflows. Thus, we expect EOdal to trigger further open-source developments, either through the release of additional Python packages that build on the existing functionalities, or through its integration into existing data processing frameworks. Ultimately, EOdal is intended to provide researchers with an alternative to proprietary software platforms while also easing the burden of data management for practitioners. For this reason EOdal is particularly well suitable for educational activities in the field of EO. This supports the transition to reproducible science, benefiting the EO community as a whole.

#### CRedit authorship contribution statement

**Lukas Valentin Graf:** Conceptualization, Software, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Gregor Perich:** Conceptualization, Software, Writing – original draft, Writing – review & editing. **Helge Aasen:** Supervision, Writing – original draft, Writing – review & editing.

#### Declaration of competing interest

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.compag.2022.107487>. Lukas Valentin Graf reports financial support was provided by Swiss National Science Foundation.

#### Data availability

Code required to rerun the analysis presented in this paper is available under: <https://doi.org/10.5281/zenodo.7278252>.

#### Acknowledgments

The authors thank Achim Walter from the Crop Science group at ETH for providing the IT infrastructure and in particular Norbert Kirchgessner for his support with data storage and implementation. Furthermore, we thank Alfred Burri for on-site support and contributions to the usage example and Fabio Oriani for valuable comments on the manuscript. The development of EOdal was conducted within the project ‘PhenomEn’ funded by the Swiss Science Foundation (grant number IZCOZO\_198091). GP acknowledges funding through the project ‘DeepField’ of the Swiss Federal Office for Agriculture (BLW).

#### References

- Asam, S., Callegari, M., Matiu, M., Fiore, G., De Gregorio, L., Jacob, A., Menzel, A., Zebisch, M., Notarnicola, C., 2018. Relationship between Spatiotemporal Variations of Climate, Snow Cover and Plant Phenology over the Alps—An Earth Observation-Based Analysis. *Remote Sens.* 10 (11), 1757. <http://dx.doi.org/10.3390/rs10111757>, URL: <http://www.mdpi.com/2072-4292/10/11/1757>.
- Barker, J.B., Heeren, D.M., Neale, C.M.U., Rudnick, D.R., 2018. Evaluation of variable rate irrigation using a remote-sensing-based model. *Agricult. Water Manag.* 203, 63–74. <http://dx.doi.org/10.1016/j.agwat.2018.02.022>, URL: <https://www.sciencedirect.com/science/article/pii/S0378377418301082>.
- Caparros-Santiago, J.A., Rodriguez-Galiano, V., Dash, J., 2021. Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. *ISPRS J. Photogramm. Remote Sens.* 171, 330–347. <http://dx.doi.org/10.1016/j.isprsjprs.2020.11.019>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0924271620303269>.
- Duarte, L., Teodoro, A.C., Monteiro, A.T., Cunha, M., Gonçalves, H., 2018. QPhenoMetrics: An open source software application to assess vegetation phenology metrics. *Comput. Electron. Agric.* 148, 82–94. <http://dx.doi.org/10.1016/j.compag.2018.03.007>, URL: <https://www.sciencedirect.com/science/article/pii/S0168169916302162>.
- Egli, M., Gärtner, H., Rösli, C., Seibert, J., Wiesenberger, G.L.B., Wingate, V., 2020. Landschaftsdynamik im Gebiet des Grossen Mooses - Moorböden, Wassermanagement und landwirtschaftliche Nutzung im Spannungsfeld zwischen Produktivität und Nachhaltigkeit. *Schriftenreihe Physische Geographie* 68, <http://dx.doi.org/10.5167/uzh-190633>, URL: <https://www.zora.uzh.ch/id/eprint/190633/>, ISBN: 9783906894164 Num Pages: 91 Place: Zürich Publisher: Geographisches Institut der Universität Zürich.
- Frantz, D., 2019. FORCE—Landsat + sentinel-2 analysis ready data and beyond. *Remote Sens.* 11 (9), 1124. <http://dx.doi.org/10.3390/rs11091124>, URL: <https://www.mdpi.com/2072-4292/11/9/1124>, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.
- Godara, S., Toshniwal, D., Parsad, R., Bana, R.S., Singh, D., Bedi, J., Jhahria, A., Dabas, J.P.S., Marwaha, S., 2022. AgriMine: A Deep learning integrated spatio-temporal analytics framework for diagnosing nationwide agricultural issues using farmers’ helpline data. *Comput. Electron. Agric.* 201, 107308. <http://dx.doi.org/10.1016/j.compag.2022.107308>, URL: <https://www.sciencedirect.com/science/article/pii/S0168169922006172>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <http://dx.doi.org/10.1016/j.rse.2017.06.031>, URL: <https://www.sciencedirect.com/science/article/pii/S0034425717302900>.
- Graf, L.V., Gorroño, J., Hueni, A., Walter, A., Aasen, H., 2022. Propagating the sentinel-2 top-of-atmosphere radiometric uncertainty into land surface phenology metrics using a Monte Carlo framework. (Submitted for Publication).

- ISO, 2015. ISO 19109:2005 Geographic information — Rules for application schema. URL: <https://www.iso.org/standard/39891.html>.
- Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *J. Hydrol.* 586, 124905. <http://dx.doi.org/10.1016/j.jhydrol.2020.124905>, URL: <https://www.sciencedirect.com/science/article/pii/S0022169420303656>.
- Marshall, M., Tu, K., Brown, J., 2018. Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. *Remote Sens. Environ.* 217, 258–271. <http://dx.doi.org/10.1016/j.rse.2018.08.001>, URL: <https://www.sciencedirect.com/science/article/pii/S0034425718303663>.
- Mittermayer, M., Maidl, F.-X., Nätscher, L., Hülsbergen, K.-J., 2022. Analysis of site-specific N balances in heterogeneous croplands using digital methods. *Eur. J. Agron.* 133, 126442. <http://dx.doi.org/10.1016/j.eja.2021.126442>, URL: <https://www.sciencedirect.com/science/article/pii/S1161030121002136>.
- Pan, S., Zhao, X., Han, Y., Wang, X., Yang, F., Wang, S., Liu, Z., Zhang, Q., Zhang, Q., Wang, K., 2022. Online information platform for the management of national variety test of major crops in China: Design, development, and applications. *Comput. Electron. Agric.* 201, 107292. <http://dx.doi.org/10.1016/j.compag.2022.107292>, URL: <https://www.sciencedirect.com/science/article/pii/S0168169922006044>.
- Perich, G., Turkoglu, M.O., Graf, L.V., Wegner, J.D., Aasen, H., Walter, A., Liebisch, F., 2022. Pixel-based crop yield mapping and prediction using spectral indices and neural networks on Sentinel-2 time series data. (Submitted for Publication).
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48 (2), 119–126. [http://dx.doi.org/10.1016/0034-4257\(94\)90134-1](http://dx.doi.org/10.1016/0034-4257(94)90134-1), URL: <https://www.sciencedirect.com/science/article/pii/0034425794901341>.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. <http://dx.doi.org/10.1016/j.rse.2019.111402>, URL: <https://www.sciencedirect.com/science/article/pii/S0034425719304213>.
- Xin, D., Miao, H., Parameswaran, A., Polyzotis, N., 2021. Production machine learning pipelines: Empirical analysis and optimization opportunities. In: Proceedings of the 2021 International Conference on Management of Data. ACM, Virtual Event China, pp. 2639–2652. <http://dx.doi.org/10.1145/3448016.3457566>, URL: <https://dl.acm.org/doi/10.1145/3448016.3457566>.