

Handling of Big Data in Agricultural Remote Sensing



Yanbo Huang¹ and Zhongxin Chen²

¹USDA Agricultural Research Services, Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS, USA

²Digitalization and Informatics Division, Food and Agriculture Organization, United Nations, Rome, Italy

Keywords

Agriculture · Remote sensing · Big data

Definition

- Agricultural remote sensing is an important technology for modern precision agriculture to handle within-field variability for site-specific management compared with uniform management in traditional agriculture.
- Big data refer to data resources that are so large in volume, fast in pattern change, or complex in structure and relationship. So, it is difficult or impossible to process big data using traditional methods. In agricultural remote sensing, frequently big data with massive volume and complexity are handled.

Overview

The acquisition, processing, storage, analysis, visualization, and application of agricultural remote sensing big data are critical to guide the success of precision agriculture. This entry overviews available data resources for agricultural remote sensing, recent development of technologies for remote sensing big data management, and remote sensing data processing and management for precision agriculture.

Remote sensing technology has been developed for earth observation (EO) with massive remotely sensed data for various research needs and practical applications. As of September 1, 2021, there are 4550 active man-made satellites in space orbiting earth and 22% of the satellites serve for EO remote sensing (DEWEsoft 2022). It can be predicted that this number will increase rapidly, especially for those satellites that will be launched into low orbits around the earth for near earth observation. For remote sensing, satellites are equipped with one or more sensors or instruments to collect various observation data of earth surface, including land, water, and atmosphere. These satellite sensors acquire images of earth surface unceasingly at different spatial, spectral, and temporal resolutions so that huge volume of remotely sensed images is available in many countries and international agencies with the data volume growing every day, every hour, and even every second.

Sustainable development of agriculture is the goal for global food security. Modern agricultural development is the result from the development of precision agriculture. Precision agriculture began to renovate agricultural operations in the 1980s when the theory and practice were established based on agricultural mechanization through the integration of global positioning system (GPS), geographic information system (GIS), and remote sensing technologies (Zhang et al. 2002). Over the past four years, precision agriculture has evolved from strategic monitoring using satellite imagery for regional decision-making to tactical monitoring and control prescribed by the information from low-altitude remotely sensed data by unmanned aerial vehicles (UAVs) for field-scale site-specific management to deal with within-field variability of soil properties, weed distributions, crop stress of insects, diseases, nutrients and water, biomass, and yield. Now data science and big data technology are gradually merged into precision agricultural schemes so that the data can be analyzed rapidly in time for decision-making although research remains for how to manipulate big data and convert the big data to “small” data and actionable information for specific issues or fields for accurate precision agricultural operation. Agricultural remote sensing is a key technology that, with global positioning data, produces spatially varied data and information for agricultural planning and prescription for precision agricultural operations with GIS (Yao and Huang 2013). Agricultural remote sensing data appear in different forms and are acquired from different sensors and at different intervals and scales. Agricultural remote sensing data all have characteristics of big data in complexity and volume. The acquisition, processing, storage, analysis, and visualization of agricultural remote sensing big data are critical to the success of precision agriculture. With the most recent and coming advances of information and electronics technologies and remote sensing big data support, precision agriculture will be developed into the stage of smart, intelligent agriculture.

Big Data in Agricultural Remote Sensing

Remote sensing technology has been developed for EO from different optical sensors and platforms. Optical sensors are mainly broad-band and narrow-band multispectral and hyperspectral for EO imaging and nonimaging spectral data acquisition. Platforms are space-borne for satellite-based sensors, airborne for sensors on manned and unmanned airplanes, and ground-based for field on-the-go proximal sensing sensors. Objects on the earth continuously transmit, reflect, and absorb electromagnetic waves. Remote sensing technology detects the objects by determining the difference of the transmitted, reflected, and absorbed electromagnetic waves. Depending on the sensors, optical remote sensing works over the electromagnetic spectrum at the spectral bands of visible (400–750 nm), including red, green, and blue bands (RGB), infrared (750 nm – 25 μ m), including near-infrared (NIR), shortwave infrared (SWIR), and thermal infrared (TIR), and microwave (25 μ m – 1 mm). All these factors with geospatial distribution and data acquisition frequency result in remote sensing big data with huge volume and high complexity.

Remote sensing technology has been developing for innovatively high-performance sensors used for high spatial, spectral, and temporal resolution data. Agricultural remote sensing is a specialized field of research and application with images and spectral data in huge volume and high complexity for decision support for agricultural ecosystem assessment and production management. Over the years, remote sensing has been conducted for analyzing soil properties, mapping crop varieties and cropping practices, monitoring crop growth and detect crop stress for decision support in crop fertilization, irrigation, pest management, crop insurance, and food production. Typically, agricultural remote sensing systems use visible-NIR sensors for crop growth studies, SWIR sensors for soil and plant moisture studies, TIR sensors for crop canopy temperature studies, and microwave sensors for soil moisture studies (Moran et al. 1997; Pinter et al. 2003;

Mulla 2013). Besides, new sensors such as LiDAR (Light Detection and Ranging) and SAR (Synthetic Aperture Radar) sensors have been applied to measure vegetation structure and soil properties over agricultural crop lands.

With the rapid development of remote sensing technology, especially the use of new sensors with high resolutions, the volume of remote sensing data is dramatically increasing with a high complexity.

Remote Sensing Data Processing and Products

Remote sensing acquires images in general as raw data. The raw data must be further corrected due to artifacts, attenuations, and deformations from interactions between sensors, atmospheric conditions, and terrain profiles on the earth surface. The corrections typically include radiometric and geometric corrections. The radiometric correction is related to the sensitivity of the remote sensor, topography and sun angle, and atmospheric scattering and absorption. The atmospheric correction is important for agronomic applications, especially when satellite data are used. For atmospheric correction, the data of atmospheric conditions during image acquisition are needed. However, the data conditions typically vary with time and location. The geometric correction fixes squeezing, twisting, stretching, and shifting of remotely sensed image pixels relative to the actual position on the ground, which are caused by remote sensing platform's angle, altitude and speed, sensitivity of the remote sensor, and earth surface topography and sun angle, atmospheric scattering and absorption, and rotation of the earth. The raw and corrected remote sensing images can be categorized into data products at different processing levels for use (Table 1). The programs of MODIS (moderate resolution imaging spectroradiometer) (National Aeronautics and Space Administration (NASA), Washington, D.C.), Landsat (NASA and United States Geological Survey (USGS), Reston, VA), European Space Agency (ESA) Copernicus program satellites such as Sentinel series (ESA ESRIN,

Frascati, Italy), and Chinese satellites such as Ziyuan (China Centre for Resources Satellite Data and Application, Beijing, China) all provide products at different levels with different applications. Thus, a remote sensing image can be corrected and processed into various products at different levels and the same image product could be resampled to scale up or down to meet the practical requirements and transformed for different applications so that the volume and complexity of remote sensing data are rapidly increased as big data.

Remote sensing began with aerial photography in early nineteenth century. Satellite remote sensing has been developed and dominating since the 1960s. In the last decade, airborne remote sensing, especially UAV-based remote sensing, has been significantly developed and applied for monitoring natural resources and agricultural lands. With the advancement of remote sensing systems and methods, the new remote sensing data product system can be formulated for airborne remote sensing data processing flow. UAV remote sensing systems are often operated at very low altitude below clouds, especially for precision agriculture for very high-resolution images (1 mm–5 cm/pixel). So, the images from UAV systems could be simplified to ignore signal attenuation without atmospheric correction. A UAV image may only cover a small part of the field. To completely cover the field, mosaicking of sequential UAV remote sensing images is required through processing of the raw images. The processing of the sequential

Handling of Big Data in Agricultural Remote Sensing, Table 1 Remote sensing data products

Processing level	Description
0	Raw data
1	Radiometrically and geometrically corrected and geo-located data represented in a map projection
2, 3, and 4	Derived agro-geophysical variables at the same resolution and location as level 1 source data, variables mapped on uniform spatial grid scales, and model output or results from analyses of lower-level data

image data also provides three dimensional models of the ground surface with structure from motion (SfM) point clouds using stereo vision for extraction of surface features such as plant height and biomass volume (Bendig et al. 2014; Huang et al. 2017).

Remote Sensing Data Analysis

Analysis of remote sensing data can be qualitative or quantitative based on purposes of the analysis. Qualitative remote sensing data analysis is classification-based and critical to transform remote sensing data into useful information for practical applications. Remote sensing image classification is based on image pixel classification in unsupervised and supervised modes. The commonly used methods include ISODATA (Iterative Self-Organizing Data), maximum likelihood, and machine learning algorithms such as artificial neural networks and support vector machine. With the development of remote sensing technology for high-resolution data, the pixel-based classification methods cannot classify the high-resolution images with more details clustered into some unknown “blank” spots. Object-based remote sensing image classification provides a solution to perform image segmentation to merge the neighboring pixels with similar spectral signatures into objects as “pixels” to classify. Quantitative remote sensing data analysis is model based to empirically model the features, such as vegetation indices, extracted from remote sensing data, with biophysical and biochemical measurements, such as plant height, shoot dry weight, and chlorophyll content. With the calibrated models, the biophysical and biochemical parameters can be predicted for estimation of biomass amount and crop yield.

The quantitative remote sensing data analysis can also develop physically based model based on the underlying theory of radiative transfer to conduct spectra simulation and plant biophysical and biochemical parameter inverse to understand the mechanism of interaction between remote sensing and ground surface features. The

PROSPECT leaf optical properties model (Jacquemoud and Baret 1990) and the SAIL canopy bidirectional reflectance model (Verhoef 1984) are two representatives widely used in radiative transfer studies of remote sensing plant characterization.

Although there are many remote sensing satellites and sensors to collect land surface images for agricultural big data, sometimes the data is still inadequate for timely information service for agricultural practice and management in terms of quantitative data items and temporal frequency. For this purpose, various data assimilation techniques have been developed to integrate process-based models (such as crop growth model, soil hydrological model) and in situ observations with time-series remote sensing imagery to fill the data gaps and to generate more quantitative agronomic data items, which are ready to be used in precision agriculture.

In recent years, machine learning and deep learning have been developed for remote sensing image classification. Deep learning was believed to be crucial and important in advancing remote sensing data processing and analysis, particularly for handling of remote sensing big data. With the low-level spectral and textural features in the bottom level, the deep feature representation in the top level of the deep artificial neural network can be directly fed into a subsequent classifier for pixel-based classification. This hierarchy handles deep feature extraction with remote sensing big data, which can be used in all parts in remote sensing data processing and analysis.

Remote Sensing Data Visualization and Application

After data acquisition and correction, big data is processed and qualitatively or quantitatively analyzed, and the resulted data and information are usually visualized and applied in a geospatial tool, such as Geographical Information System (GIS). Visualization of remote sensing data and products are critical for data interpretation and analysis.

GIS as a platform of remote sensing data visualization has been developing in four aspects:

- **Modularization**
Modular GIS is organized of components with certain standards and protocols.
- **Web enabling**
Web GIS has been developed to publish geospatial data for viewing, query, and analysis through Internet.
- **Miniaturization and mobility**
Although desktop GIS applications still dominate, mobile GIS clients have been adopted with tablets and smart phones.
- **Data-based and interoperability.**
GIS spatial data management has been developed from flat file management, file/database management, to spatial database management. Spatial data management provides the capabilities of massive data management, multiuser cocurrent operation, data visit permission management, and cocurrent visit and systematic applicability of database clusters. Big data from different sources is organized and managed based on data standards (e.g., ISO and OGC standards) to enable the data interoperability.

The integration of remote sensing data with GIS has been developed in the past two decades. Techniques such as machine learning and deep learning offer great potential for better extraction of geographical information from remote sensing data and images. However, issues remain as data organization, algorithm construction, and error and uncertainty handling. With the increased volume and complexity of remote sensing data acquired from multiple sensors using multispectral and hyperspectral devices with multiangle views with the time, new development is needed for visualization tools with spatial, spectral, and temporal analysis.

Big Data Management in Remote Sensing

The generalization, standardization, and serialization of remote sensing data products are crucial for reliable and consistent remote sensing data applications from multiple sources. In designing the structure of remote sensing, data management for image mapping on the 3D sphere is a central issue. To solve the problem, the earth surface can be virtually divided into blocks. In each block, the pyramid of images is created, and the scene can be visualized with image blocks at different resolutions with the altitude of viewing point. World Wind (NASA), GoogleEarth (Google Inc., Mountain View, CA), and BingMaps (Microsoft, Redmond, WA) use this method to visualize geospatial data. However, the method at higher resolution will lead to the problem of computing floating numbers of the degree of blocking. Manipulation of floating numbers may cause problems in data processing of remote sensing images to lose computing precision significantly, and at the same time, it may cause inaccurate mapping of sphere texture. In addition, the image size of the method with different block sizes at different resolutions cannot match with commonly used map scales.

A blocking/tile remote sensing data analysis and access structure called Data Cube has been established to analyze and publish Landsat data Geoscience Australia archived covering the Australian continent (<http://www.datacube.org.au/>). The Data Cube has made more than three decades of satellite imagery spanning Australia's total land area become available and provides over 240,000 images showing how Australia's vegetation, land use, water movements, and urban expansion have changed over the past 30 years.

Scientists of Chinese Academy of Science developed an innovative five-layer-fifteen-level (FLFL) structure for managing satellite imagery at different resolutions (Wang et al. 2012). This data management structure is well fitted for remote sensing data organization with the grid of latitude and longitude to match up with the commonly used map scales. For high-resolution

satellites, manned aircrafts, and UAVs for low-altitude remote sensing, the FLTL structure is extended into a four-layer-twelve-level (FLTL) structure (Huang et al. 2018). The FLTL structure is used to block the images on the sphere surface of the earth and the images are from high-resolution satellite sensors, manned airborne sensors, and UAV-based sensors with the pixel resolution from 5.57 m to 1 mm. Any images in coarser or finer resolutions can be further handled with expansion of the high-resolution and low-resolution ends of the FLTL structure. With the FLTL structure, at the end of the highest resolution the pixel resolution is 1 mm, with which a 1:1.5 image map can be made.

Applications of Agricultural Remote Sensing Big Data

UAV Remote Sensing for Crop Science Studies in a Research Farm in Mississippi

A research farm within the campus of Mississippi State University (MSU) (33.472069°N, 88.773452°W) has been focused with professors of MSU and scientists of the US Department of Agriculture, Agricultural Research Services for various research studies of crop science (Fig. 1). Within the area of about 150 hectares over this farm, remote sensing monitoring has been conducted. Over the past few years, remote sensing data have been accumulated with for various studies of soil, plant, and irrigation, which are mainly UAV-based images of RGB, multispectral, TIR, hyperspectral, and LiDAR over the fields as shown in Fig. 1. In image processing and analysis, various methods and algorithms have been being developed and applied on standalone personal computing workstations and high-performance computing platforms, such as a Cray CS500 Linux Cluster with 11,520 2.40GHz Xeon Platinum 8260 processor cores created at MSU (<https://www.hpc.msstate.edu/computing/atlas/>), and with the increase of the data volume and complexity, the challenges at data storage, computation, and system input and output (I/O) in data management and application will be issues to solve.

The adoption of the FLTL structure for UAV remote sensing data management can help distributed data storage, computational decomposition for parallel processing, and relief of limited system I/O capability, which would result in effective applications driven by processing and analysis of the data with such management. Figure 2 shows UAV imaging points using MicaSense Altum camera (MicaSense, Seattle, WA) which integrates a radiometric thermal imager with imagers of five discrete spectral bands, including blue, green, red, red-edge, and NIR, allowing various plant science studies in field. Figure 3 shows the blocking and resampling scheme of the orthomosaicked color infrared composite image of the flyover area of the research farm with the FLTL scheme in the blocks of $0.0005^\circ \times 0.0005^\circ$ at the seventh level in the third layer. Figure 4 shows the blocking and resampling scheme of the orthomosaicked thermal image of the flyover area of the research farm with the FLTL scheme in the blocks of $0.01^\circ \times 0.01^\circ$ at the third level in the first layer. The raw multispectral and thermal images were acquired using the Micasense Altum camera mounted on a DJI Inspire 2 quadcopter UAV (DJI, Shenzhen, China). With the flight altitude of 120 m, the camera provided the spatial resolutions of 5.6 cm/pixel and 81 m/pixel, respectively, for multispectral images and thermal images. Multiple images were acquired to be mosaicked to cover the farm area using Pix4D Mapper software (Pix4D S. A., Prilly, Switzerland) and the resulting orthomosaic images were georectified and resampled to fit into the FLTL structure.

FAO Hand in Hand Geospatial Platform

Food and Agriculture Organization (FAO) of the United Nations is leading international efforts to fight hunger and poverty and promote information and knowledge sharing in the field of food and agriculture. In the recent years, FAO has a lot of projects using remote sensing big data to facilitate the implementation of UNSGDs, especially SDG1 (No poverty) and SDG2 (Zero hunger), and digital transformation of agriculture. FAO Hand-in-Hand Geospatial Platform (HiH GP, <https://www.fao.org/hih-geospatial-platform/en/>)



Handling of Big Data in Agricultural Remote Sensing, Fig. 1 PlanetScope (Planet, San Francisco, CA) satellite image scene of the research farm within the

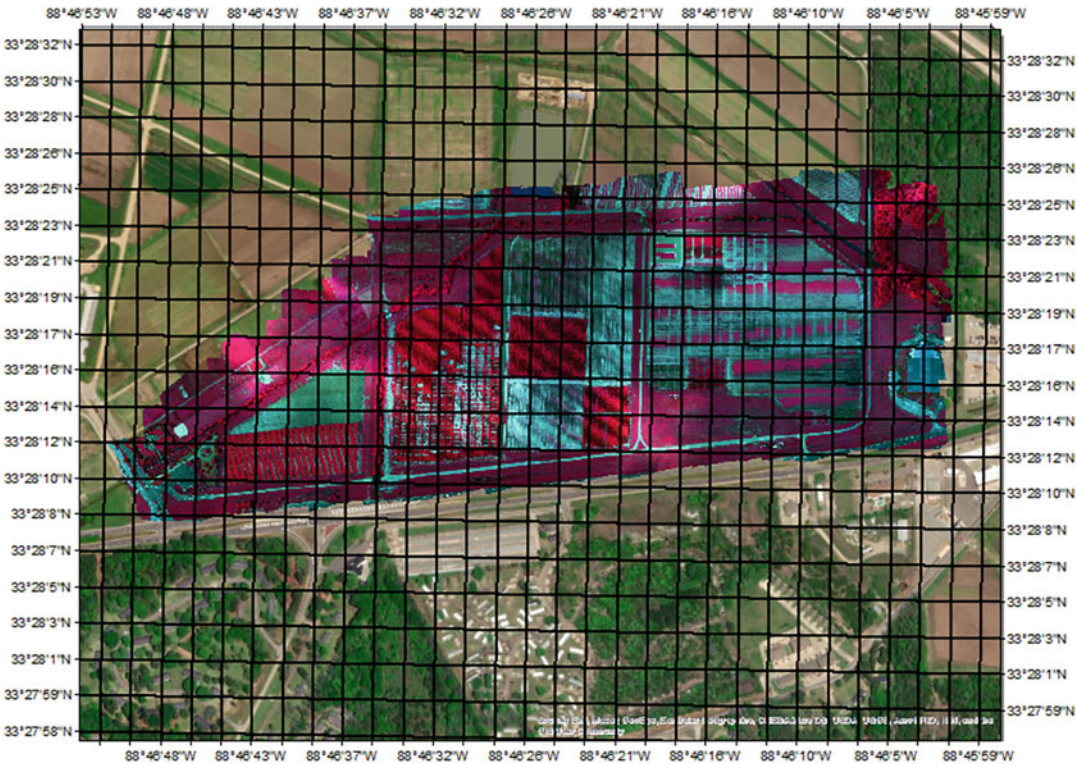
campus of Mississippi State University and focus area of UAV flyovers (within red-line polygon)

is the FAO geospatial big data platform which gives access to majority of FAO's geospatial data, including agricultural remote sensing big data. HiH GP gives easy access to over 4000 data series combining over millions of geospatial layers. The platform uses the cutting-edge information technologies available, including advanced geospatial modeling and analytics, to identify and maximize the opportunities to raise incomes, and reduce inequalities in agricultural production and rural development. The platform provides an evidence-based view of economic opportunities and aims at improving the targeting and tailoring of policy interventions, driving innovation, encouraging finance and investment flows, and institutional reforms. HiH GP was originally developed for the big data analysis to the UN

initiative of Hand-in-Hand. After its launching in 2020, its services are not confined to HIH, but open to all of users in the world. It acts as the FAO geospatial infrastructure for data integration, federation, interoperability, knowledge discovery, and information sharing and becomes digital public goods in the domain of agriculture, environment, and sustainability. The integration of FAO's broad thematic data sets (trade, crops, livestock, fishery, forestry, land, energy, water, climate, and much more) facilitates the platform with the most up-to-date food and agriculture-related information. FAO HiH GP is a good example to show the concept of agricultural remote sensing big data, including data acquisition, processing, storage, analysis, visualization, and application (Fig. 5).

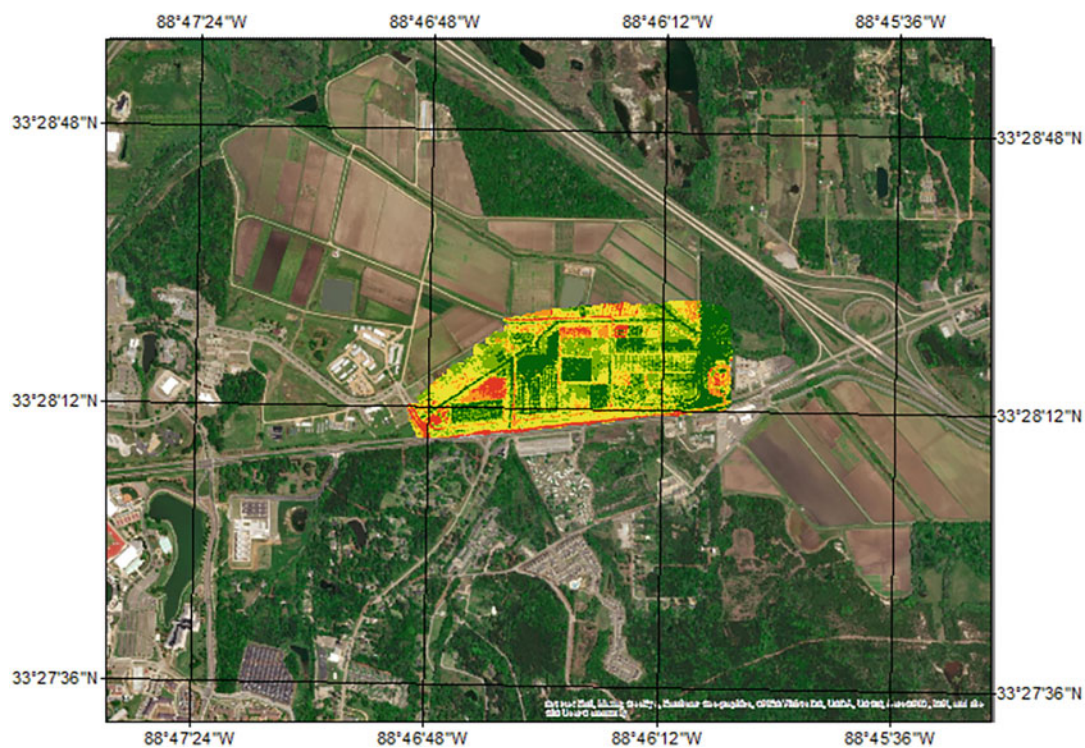


Handling of Big Data in Agricultural Remote Sensing, Fig. 2 UAV imaging points using MicaSense Altum camera

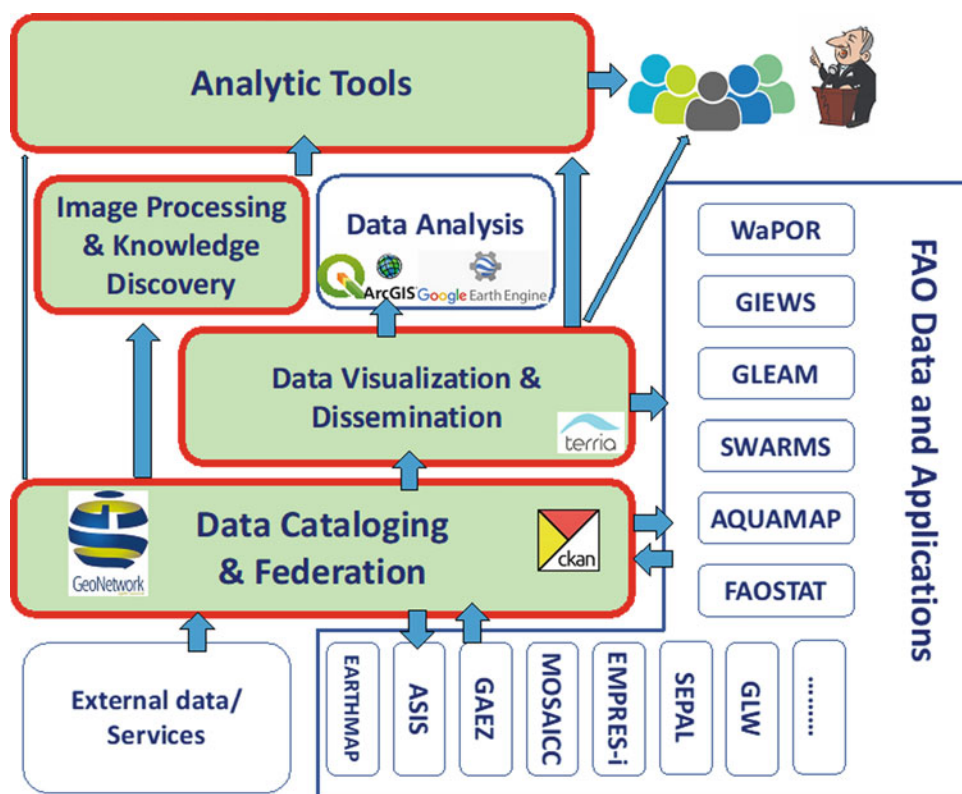


Handling of Big Data in Agricultural Remote Sensing, Fig. 3 Blocking and resampling scheme of the orthomosaicked UAV color-infrared composite image of the

research farm area of MSU with the FLTL scheme in the blocks of $0.0005^{\circ} \times 0.0005^{\circ}$



Handling of Big Data in Agricultural Remote Sensing, Fig. 4 Blocking and resampling scheme of the orthomosaicked UAV thermal image of the research farm area of MSU with the FLTL scheme in the blocks of $0.01^\circ \times 0.01^\circ$



Handling of Big Data in Agricultural Remote Sensing, Fig. 5 Technical framework of FAO HiH-GP for processing and federating remote sensing big data in agriculture

References

- Bendig J, Bolten A, Bennertz S, Broscheit J, Eichfuss S, Bareth G (2014) Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote Sens* 6:10395–10412
- DEWEsoft (2022) Every satellite orbiting earth and who owns them. <https://dewesoft.com/daq/every-satellite-orbiting-earth-and-who-owns-them>. Accessed 25 Apr 2022
- Huang Y, Brand H, Sui R, Thomson SJ, Furukawa T, Ebelhar MW (2017) Cotton yield estimation using very high-resolution digital images acquired on a low-cost small unmanned aerial vehicle. *Trans ASABE* 59: 1563–1574
- Huang Y, Chen Z, Yu T, Huang X, Gu X (2018) Agricultural remote sensing big data: management and applications. *J Integr Agric* 17(9):1915–1931
- Jacquemoud S, Baret F (1990) PROSPECT: a model of leaf optical properties spectra. *Remote Sens Environ* 34: 75–91
- Moran MS, Inoue Y, Barnes EM (1997) Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens Environ* 61:319–346
- Mulla DJ (2013) Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosyst Eng* 114:358–371
- Pinter PJ, Hatfield JL Jr, Schepers JS, Barnes EM, Moran MS, Daughtry CST, Upchurch DR (2003) Remote sensing for crop management. *Photogramm Eng Remote Sens* 69:647–664
- Verhoef W (1984) Light scattering by leaf layers with application to canopy reflectance model: the SAIL model. *Remote Sens Environ* 16:125–141
- Wang D, Zheng F, Lai J, Yu T, Li J, Guo S (2012) A new parallel algorithm based on five-layer fifteen-level remote sensing data organization. *Microcomput Informat* 1:1–5
- Yao H, Huang Y (2013) Remote sensing applications to precision farming. Chap. 18. In: Wang G, Weng Q (eds) *Remote sensing of natural resources*. CRC Press, Boca Raton, pp 333–352
- Zhang N, Wang M, Wang N (2002) Precision agriculture—a worldwide overview. *Comput Electron Agric* 36(2–3):113–132