DIVISION OF AGRICULTURE
RESEARCH & EXTENSION
University of Arkansas System

Agriculture and Natural Resources

FSA2216

Spatial Data Types and Properties

Ujjwal Sigdel
Graduate Research Assistant - Crop, Soil and Environmental Sciences

Aurelie Poncet
Assistant Professor of Precision Agriculture - Crop, Soil and Environmental Sciences

Jason A. Davis
Assistant Professor and State Extension Specialist of Remote Sensing and Pesticide Application - Crop, Soil and Environmental Sciences

Wesley France
Program Associate - Crop, Soil and Environmental Sciences

Arkansas Is Our Campus

Visit our website at: https://www.uaex.uada.edu

Introduction

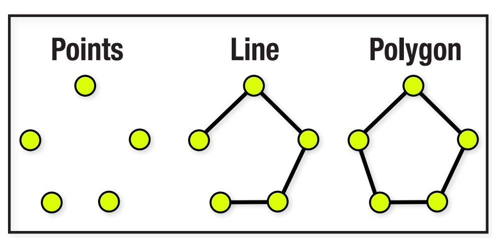
Spatial data contains information about a geographic location or area and can be represented in vector or raster form (Winter, 1998). Both vector and raster data are used in agronomy, environmental sciences, and engineering among other disciplines to understand site-specific processes occurring within an area and inform decision-making. This factsheet is intended to define and provide context for spatial data types, formats, and usage.

Spatial Data Types

Vector data may be represented as points, lines, or polygons (Figure 1). Points are zero-dimensional objects with a spatial reference frame. Lines are one-dimensional objects that connect points into a path. Polygons are two-dimensional objects connecting a path into an area. The points that define the shape of line and polygon vector data are referred to as vertices. Vector data are used to represent discrete features with well-defined boundaries such as rivers, transportation routes, buildings, or places where specific events may have occurred. In production agriculture, vector data sets are commonly used to represent field boundaries, traffic or scouting paths, irrigation or guidance lines, and soil or tissue sampling locations (Figure 2a).

Raster data consists of pixels organized in a grid. They are used to represent one or multiple layers of spatially continuous data such as air temperature, rainfall amounts, elevation, or surface radiometric reflectance. There are three configurations of raster data. **Configuration 1:** A single-band raster depicts one variable or layer of information. Digital surface and elevation models, which capture different aspects of terrain in a study area, are common examples of single-band raster data (Figure 2b). **Configuration 2:** A multi-band raster represents multiple variables that describe different spectral or thematic aspects of the spatial data. Each variable in a multi-band raster is stored in a separate layer, and all layers have identical

Figure 1. Spatial vector data can be represented using points, lines, or polygons.



spatial coordinates. The images collected using red, green, blue (RGB) or multispectral sensors mounted on satellites or small unmanned aircraft systems (sUAS, or drones) are examples of multi-band rasters (Figure 3). RGB sensors are used to create RGB image composites that characterize visible changes in color within a study area. Multispectral images are used to compute vegetation indices that indicate spatial changes in land use or crop health. **Configuration 3:** A raster stack is a collection of multiple single-band raster stacked together. Each layer in the stack may represent different variables and spatial coordinates. Raster stacks are commonly used to store data collected over time or using different sensors into the same data structure (Figure 4).

Figure 2. Sample a. vector and b. raster data collected in a soybean production trial conducted in Lincoln County, Arkansas. Three vector layers are shown on the left. Points, lines, and polygons are used to represent the soil sampling locations, scouting path, and field boundary, respectively. One single-band raster layer is shown on the right to represent in-field changes in field elevation. This digital elevation model has a 3.3 ft spatial resolution and was downloaded from the Arkansas Spatial Data Infrastructure, a free repository of spatial data for Arkansas (U.S. Geological Survey, 2016).

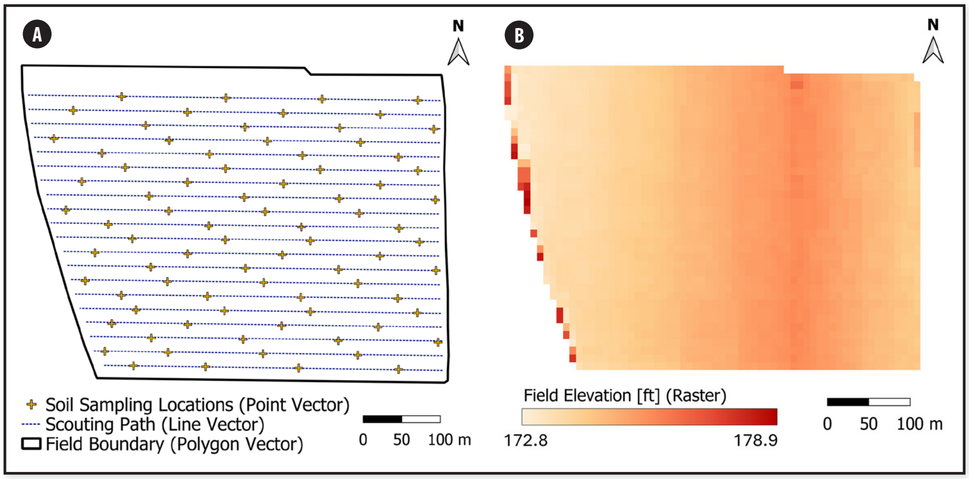
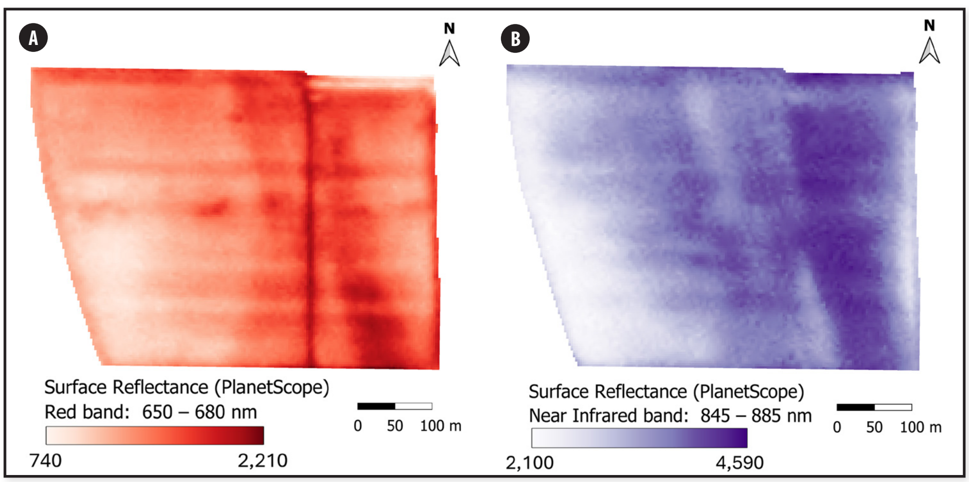


Figure 3. Sample multispectral multi-band raster data downloaded from the PlanetScope satellite image repository (Image© 2021 Planet Labs PBC). Surface reflectance in eight regions of the electromagnetic spectrum, also referred to as bands, were quantified using digital numbers and stored in the form of a multi-band raster data. The higher the digital number value, the greater the amount of radiation emitted by the surface in the corresponding band. The data collected on June 28, 2022, in the a. red and b. near infrared bands were represented using a red and purple color scale, respectively.



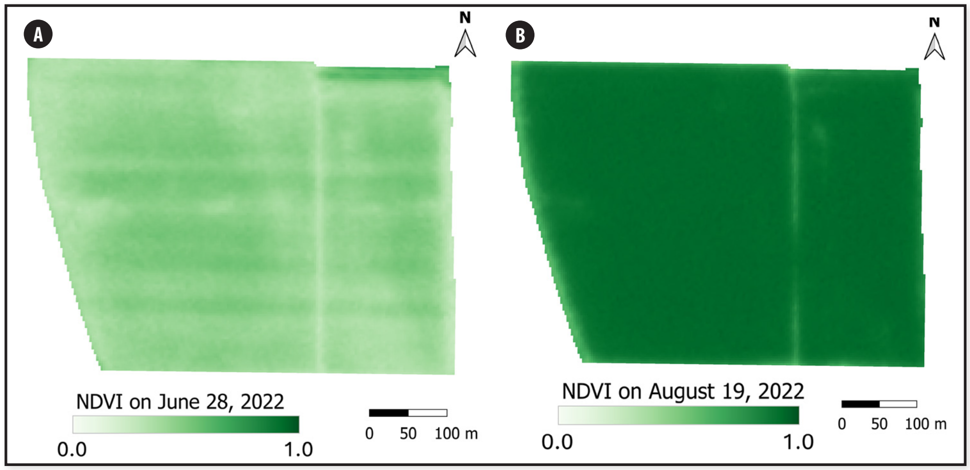
Spatial Coordinates and Extent

The physical location of spatial data on the three-dimensional surface of the Earth is uniquely identified using three numerical values known as latitude, longitude, and elevation. Latitude identifies the North-South location of a point in comparison to the Equator (0o Latitude). Longitude identifies the East-West location of a point in comparison to the Prime Meridian (0o Longitude), which is an arbitrary line running North-South passing through the Greenwich Observatory in London, UK. Elevation defines the vertical distance between the location and a surface of reference such as mean sea level (MSL).

The physical location of spatial data projected on the two-dimensional surface of a map is uniquely identified using two numerical values referred to as northing and easting. Northing and easting define the North-South and East-West position of a point on a two-dimensional map, respectively. Combinations of latitude, longitude, and elevation in the three-dimensional reference frame, or northing and easting for a two-dimensional map, define the spatial data coordinates.

The location of point vector data is defined by the spatial coordinates of its zero-dimensional elements. The location of line and polygon vector data is defined by the spatial coordinates of the vertices that comprise them as well as their organization into one-dimensional or two-dimensional objects. The location of a pixel is defined by its position in the raster image, its equivalent dimensions on the ground, and the position of the raster image in the physical space. The extent of spatial data defines the coordinates of the smallest rectangular, straight, bounding box that encompasses all spatial units (e.g., individual points, lines, polygons, or pixels, Figure 5). Spatial extent is commonly defined as the coordinates of the bounding box's four corners.

Figure 4. Sample raster stack data created from PlanetScope satellite images collected on a. June 28, 2022, and b. August 19, 2022. The normalized difference vegetation index (NDVI) was calculated as a function of the red (R) and near infrared (NIR) data collected at each date using the following equation: (NIR – R)/(NIR + R). The NDVI values are dimensionless, meaning they do not have a unit. The NDVI values range from -1 to 1, with greater values describing denser and healthier vegetation. Each NDVI image is a single-band raster. They were then grouped together into a single raster stack unit to facilitate spatio-temporal modeling.



Spatial Attributes

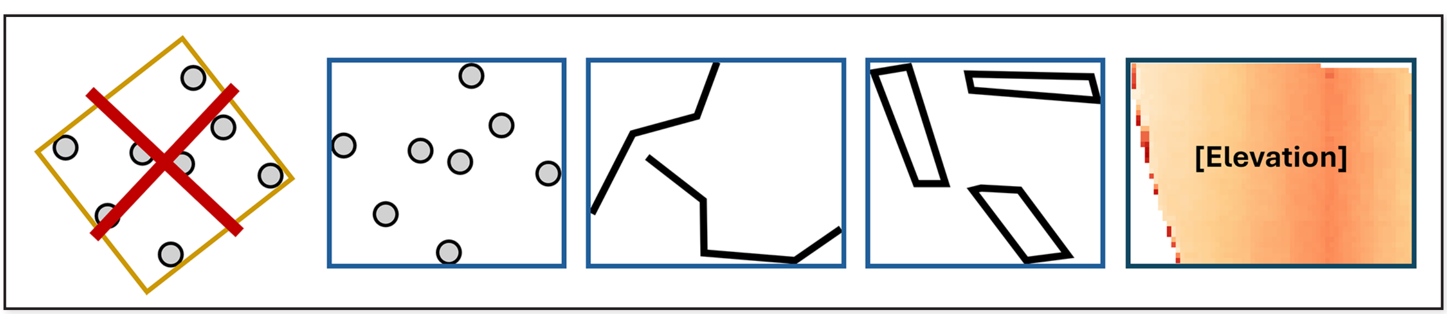
The information associated with each spatial unit is referred to as a spatial attribute. Spatial attributes can be qualitative (classes) or quantitative (numeric). For instance, county, species, and crop would provide qualitative information about a spatial unit while elevation, age, and yield would be considered quantitative attributes. More than one attribute may be associated with one vector object (Table 1). One, and only one, attribute may be associated with pixels within a raster.

Spatial Data Layers

Several points, lines, or polygons may be stored within one vector data layer, with only one type of spatial data used in each layer. Vector data layers are commonly stored in shapefiles which consist of a combination of up to eight files. The SHP and SHX files define the spatial features and locations; both are necessary. The DBF file contains the spatial data attributes. The PRJ file contains information about the reference system used to define the spatial coordinates. The XML file contains the non-spatial information associated with the entire spatial data layer, also referred to as metadata. Metadata may include information about how and when the data were collected. The CPG, SBN, and SBX files may contain additional insight into the language used to store the data and computer-readable instructions to optimize data handling and analysis.

Raster layers are defined by the number of pixels along the image height and width and the

Figure 5: Spatial extent of vector and raster data represented in blue. The bounding box represented in gold is not straight and cannot be used to define a spatial extent.



Latitude [Degrees]	Longitude [Degrees]	Elevation [ft]	Seeding Rate [seeds/ac]	Soil Texture
33.98053 N	91.70374 W	174.9	150,000	Silt Clay
33.98104 N	91.70490 W	177.1	150,000	Clay Loam
33.97243 N	91.70061 W	176.4	150,000	Clay Loam
33.97893 N	91.70764 W	173.7	125,000	Silty Clay Loam

Table 1. Sample spatial attributes associated with the sampling locations represented in Figure 2a. Latitude, longitude, and elevation describe the spatial coordinates of each sampling location. The seeding rate is a quantitative attribute that defines the number of soybean seeds applied per acre. Soil texture is a qualitative attribute that defines the soil textural class.

ground sampling distance (GSD). The GSD is a standardized measure of spatial resolution that characterizes the distance between the center of two adjacent pixels on the ground. For instance, the GSD of the image provided in Figure 2b is 15 ft. This means that the distance between the center of two adjacent pixels in that image is 15 ft. The smaller the GSD, the more pixels per unit area, and the greater the image resolution (Figures 6 and 7). Raster data layers and their metadata are stored in single file saved using one of several possible formats including TIFF, JPEG, and PNG. Vector and raster data layers may be created, managed, and analyzed using geographic information systems, or GIS (Maguire, 1991).

Figure 6. To-scale comparison of two pixels with a hypothetical 30 ft and 60 ft ground sampling distance (GSD). The difference in areas covered by two pixels with different GSDs is inversely proportional to the square of the difference in GSD. In other words, multiplying the GSD by 2 results increases the pixel area by 2x2 = 4.

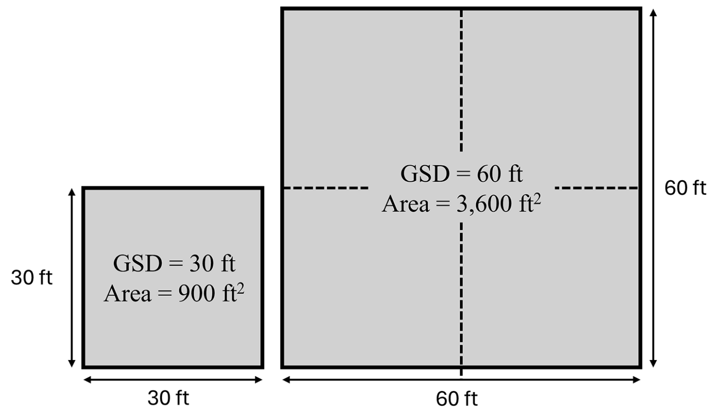
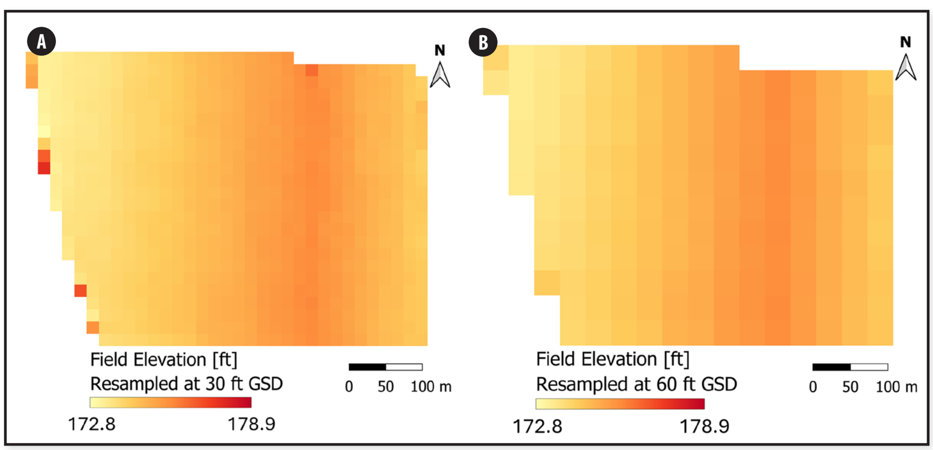


Figure 7. Resampled field elevation at a. 30 ft and b. 60 ft ground sampling resolution (GSD). Note that the higher the GSD, the smaller the number of pixels and the less detailed the data.



References

Maguire, D.J., 1991. *An Overview and Definition of GIS*. Geographical information systems: Principles and applications, 1(1), pp.9-20.

U.S. Geological Survey, 2016. 3D Elevation Program 1-Meter Resolution Digital Elevation Model. Retrieved from https://apps.nationalmap.gov/downloader/

Winter, S., 1998, November. *Bridging Vector and Raster Representation in GIS*. In Proceedings of the 6th ACM international symposium on advances in geographic information systems, pp. 57-62.

UJJWAL SIGDEL is a graduate student at the University of Arkansas in Fayetteville. **AURELIE PONCET** is an assistant professor of precision agriculture with Crop, Soil and Environmental Science Department. **JASON DAVIS** is an assistant professor of remote sensing and pesticide application with Crop, Soil and Enviromental Sciences located at the Livestock and Forestry Research Station. **WESLEY FRANCE** is a program associate with Crop, Soil and Environmental Science Department. All are with the University of Arkansas System Division of Agriculture.

FSA2216-PD-1-2025