

Introduction to Geospatial Data Analysis

Introduction to Geospatial Data Analysis

Key concepts, workflow, tools, and best practices

1. What Is Geospatial Data?

Aspect	Description
Definition	Information that is tied to a location on the Earth's surface (latitude, longitude, altitude, or a geographic identifier).
Types	• Vector – points, lines, polygons (e.g., city locations, road networks, land parcels). • Raster – grid cells/pixels (e.g., satellite imagery, digital elevation models).
Attributes	Non-spatial data attached to each feature (population, land-use class, temperature, etc.).
Coordinate Reference Systems (CRS)	Framework that defines how coordinates map to real-world locations (e.g., WGS84, NAD83, UTM zones).

2. Core Concepts

Concept	Why It Matters
Spatial Resolution	Determines the level of detail (e.g., 30m Landsat vs. 10m Sentinel2).
Spatial Extent	Geographic area covered by a dataset; influences processing time and storage.
Projection & Datum	Misaligned projections cause distortion; always re-project to a common CRS before analysis.
Topology	Relationships among features (adjacency, connectivity, containment) used in network analysis and error checking.
Spatial Autocorrelation	“Near things are more alike than distant things.” Basis for many statistical models (Moran's I, Getis-Ord G*).
Scale	The level of geographic detail; results can change dramatically when moving from local to regional scales.

Introduction to Geospatial Data Analysis

3. Typical Workflow

1. Define the Question

- Identify the spatial problem (e.g., “Where are the flood-prone neighborhoods?”).
- Determine required spatial and temporal resolution.

2. Data Acquisition

- **Vector sources:** cadastral databases, OpenStreetMap, government shapefiles.
- **Raster sources:** satellite imagery (Landsat, Sentinel), aerial photos, DEMs, climate grids.
- **APIs & Services:** WMS/WFS, Earth Engine, OGC standards.

3. Data Preparation

- **Cleaning:** remove duplicates, fix geometry errors, handle missing attributes.
- **Projection:** reproject all layers to a common CRS.
- **Clipping & Masking:** limit data to the study area.
- **Resampling / Re-tiling:** align raster resolutions and extents.

4. Exploratory Spatial Data Analysis (ESDA)

- Visualize layers (maps, histograms, box plots).
- Compute basic statistics (mean, median) and spatial statistics (Moran's I, hotspot analysis).
- Identify outliers and data gaps.

5. Analysis Techniques

- **Overlay & Intersection:** combine multiple vector layers (e.g., land use & flood zones).
- **Proximity & Buffering:** distance-based queries (e.g., 500m buffer around schools).
- **Network Analysis:** shortest path, service area, routing on road/utility networks.
- **Interpolation:** create continuous surfaces (IDW, Kriging) from point samples.
- **Raster Operations:** map algebra, classification, change detection.
- **Statistical Modeling:** spatial regression, Geographically Weighted Regression (GWR), machine-learning classifiers.

Introduction to Geospatial Data Analysis

6. Validation & Accuracy Assessment

- Compare results against ground truth or independent datasets.
- Use