# [AM-02-004] Overlay

## Abstract

Overlay operation is a critical and powerful tool in GIS that superimposes spatial and attribute information from various thematic map layers to produce new information. Overlay operations facilitate spatial analysis and modeling processes when being used with other spatial operations (e.g. buffer, dissolve, merge) to solve real-world problems. For both vector and raster data models, the input layers need to be spatially aligned precisely with each other to ensure a correct overlay operation. In general, vector overlay is geometrically and computationally complex. Some most used vector overlay operations include intersection, union, erase, and clip. Raster overlay combines multiple raster layers cell by cell through Boolean, arithmetic, or comparison operators. This article provides an overview of the fundamentals of overlay operations, how they are implemented in vector and raster data, and how suitability analysis is conducted.

*Keywords*: Boolean operators, logic, map algebra, overlay operation, raster overlay, suitability analysis, vector overlay, weighted overlay

## Author & citation

Cai, H. (2022). Overlay. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2022 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2022.1.2.

This entry is also available in an earlier edition:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Overlay. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

## Explanation

1. Definitions
2. Introduction
3. Vector Overlay
4. Raster Overlay
5. Overlay Operations in Suitability Analysis
6. Conclusions

### 1. Definitions

**Boolean algebra**: a branch of mathematics dealing with logical operations with binary variables (true or false). Boolean algebra uses the logical operators (AND, OR, NOT, XOR), to determine whether a particular condition is true or false.

**Map algebra:** a framework to analyze gridded data values through a variety of algebraic operators, originally proposed by Dana Tomlin in the 1980s.

**Multi-Criteria Decision Analysis (MCDA):** a decision-making process for complex problems when taking multiple criteria or objectives into consideration.

**Suitability analysis**: a GIS-based and multi-criteria decision-making approach to evaluate the appropriateness or preference of locations for a specific use.

## 2. Introduction

Overlay operations are the primary means of combining layers with different themes in spatial analysis. We can imagine overlay as the vertical stacking and merging of spatial layers. All the spatial layers in an overlay operation need to use the same coordinate system to ensure the features from different layers align correctly. Overlay operation is the core of spatial analysis that integrates multiple spatial entities to seek answers to numerous questions in the real world, such as “what major roads are within a specific county”, “which hospitals are located within a specific city”, and “which area experienced urbanization in the last ten years”. To fully answer these questions, overlay operations are always used in conjunction with other spatial analysis operations, such as selection and buffer, to prepare the input layers for overlay operations. Additional processes, such as merging and dissolving, are oftentimes involved to derive further outputs (Page-Tan et al. 2021, Unwin 2019).

Overlay operations are commonly used for both vector and raster representations of points, lines, and areas, such as point-in-area, line-in-area, and area-in-area overlay (Bolstad 2019). Figure 1 is an example illustrating how a point-in-area overlay is performed using vector and raster data, respectively.

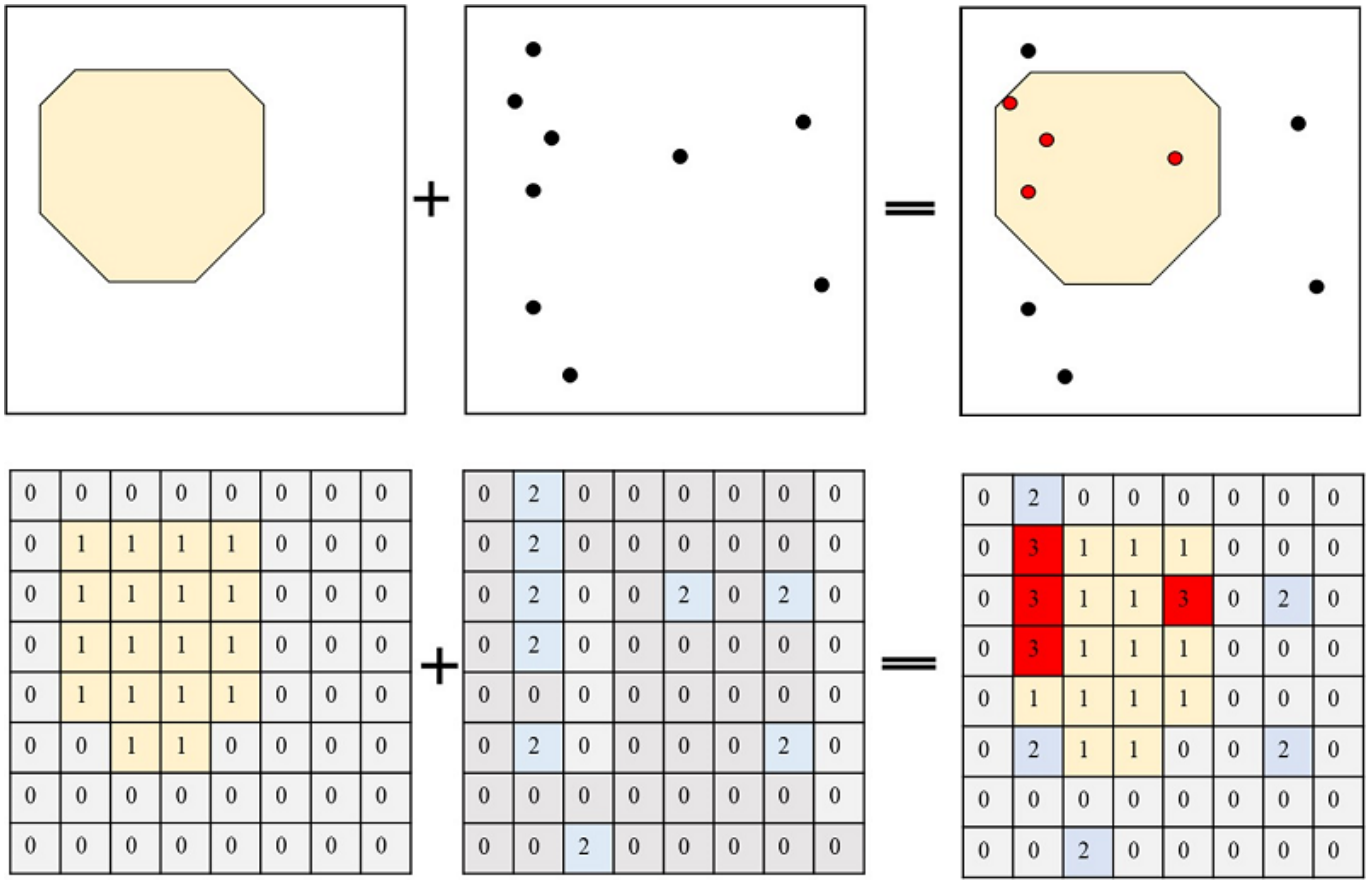


Figure 1. An example of point-in-area overlay in vector and raster data. Source: author.

Figure 1 also demonstrates that, regardless of data models, overlay analysis oftentimes involves two or more layers. Mismatched features or misaligned grids would lead to incorrect output. To prevent overlay errors, the input layers need to be properly georegistered, and referenced to the same coordinate system, map projection, datum, and resolution (raster data) whenever possible.

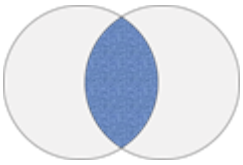
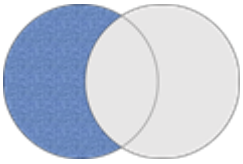
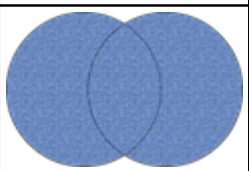
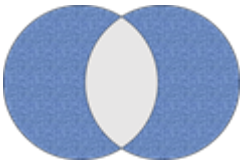
### 3. Vector Overlay

The overlay of vector data combines the point, line, and polygon features and their associated attributes from multiple data layers. Theoretically, any vector features can be overlaid with other vector feature types. Overlay operations of polygon features with any feature types (point, line, polygon) are the most frequently used in practice. The line-on-line overlay can be used to identify the intersecting points of lines. For example, we can apply a line-on-line overlay to check whether two roads intersect with each other. For roads with different Z-values (e.g. elevations), it is necessary to check the Z-value (e.g. elevation) difference between two road lines because, for example, a bridge can overpass an on-the-ground road.

Boolean algebra plays an important role in defining vector overlay operations. In GIS, it is a form of binary logical operations that links two spatial selection criteria using the Boolean operators AND, OR, NOT, and XOR. Table 1 introduces the four basic Boolean operators and

their applications. A column of the Venn diagram was used to explain the operation process. For each Boolean operator, the two circles denote two criteria, and the “true” results are highlighted in blue. In vector overlay, some commonly used operations are corresponding to the Boolean operations. For example, the intersection operation is based on the AND operator, meaning the output (true results) meets both criteria. An application can be identifying the areas which are located inside both the flood zone and developed land. A union operation corresponds to the OR operator in which the output meets at least one criterion. The union operation can be used to answer questions such as which areas are inside either the flood zone or developed land.

**Table 1. Boolean Algebra Operators**

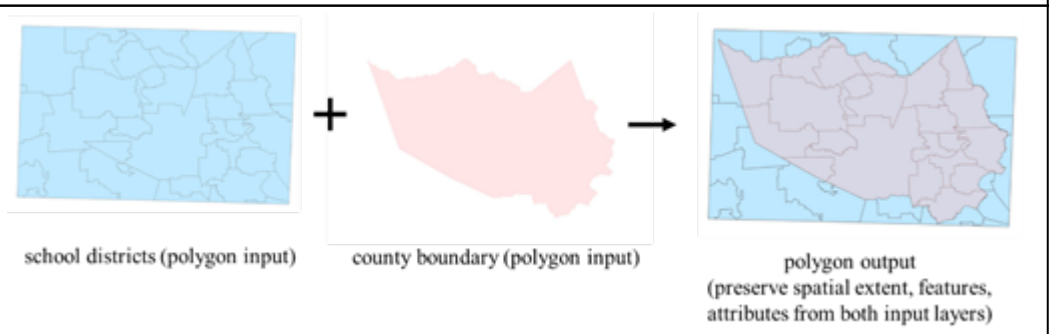
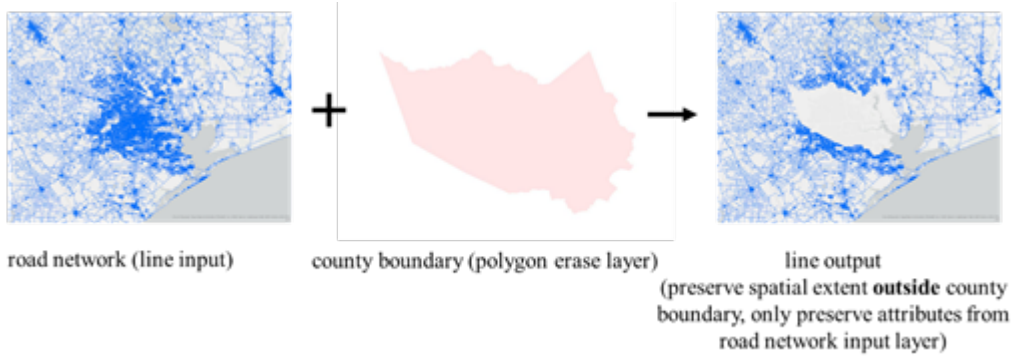
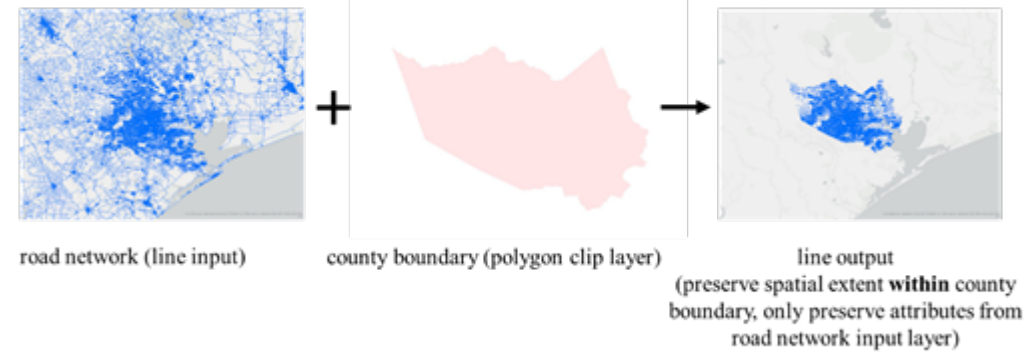
| Boolean Operator | Definition & Question Example | Venn Diagram | Corresponding Vector Overlay |
|---|---|---|---|
| AND | The true results must meet two criteria.<br>"Which areas are inside a flood zone and developed land?" |  | Intersection |
| NOT | The true results meet one criterion but not the other one.<br>"Which areas are inside a flood zone, but not developed land?" |  | Erase |
| OR | The true results meet at least one criterion.<br>"Which areas are inside either a flood zone or developed land?" |  | Union |
| XOR | The true results meet the two criteria but only when both are not true.<br>"Which areas are either inside a flood zone or developed land, but not both in the same space?" |  | Symmetrical Difference |

A comprehensive list of the common vector overlay operations includes intersection, union, clip, erase, identity, symmetrical difference, update, and split. The operations are different from each other by which feature types are allowed as the input layers, which spatial extent and attributes are preserved in the output layer. An identity operation can take an input layer in point, line, or polygon format and an identity feature in polygon or the same geometry type as another input layer. The output keeps all features in the input layer as well as the intersection of the input features and identity features. The symmetrical difference operation requires that the two input layers have the same geometry type and only features that do not overlap will be written to the output feature class. Table 2 illustrates the other four widely used vector overlay operations, intersection, union, erase, and clip, and their application examples.

The input layers of an intersection operation can be points, lines, or polygons, and the output feature class can only have the same geometry or a lower dimension as in the input layers. The output layer of an intersection preserves the attribute information from all input layers, but only contains the features in the common spatial extents of the input layers. A union operation can only be applied to polygon feature types and the output layer

preserves features and attributes information from all the input layers. As shown in the application example of union, the school districts (polygons) and the county boundary (polygon) create a new polygon output layer that keeps all the features and attributes from input layers. Erase operation is to preserve features from an input point, line, or polygon feature layer outside the spatial extent of an erase layer, whereas clip is opposite to erase, which only preserves the features inside a clip layer. Clip and intersection are similar because the output extent is defined by the common area of the input layers. However, unlike union and intersection, both erase and clip operations only retain the attribute information of the input feature to be clipped or erased.

**Table 2. Common Vector Overlay Operation Examples**

| Operation | Application Example |
|---|---|
| Intersection |  |
| Union |  |
| Erase |  |
| Clip |  |

The four vector overlay operations are powerful tools to subset or combine spatial coverage, geographic, and attribute information from multiple layers. The overlay outputs oftentimes work as inputs for other spatial operations, such as dissolve and merge, to further derive new information. For example, the intersection example in Table 2 uses flood zone areas (polygons) and county boundaries (polygons) as input layers and yields an intermediate output layer with new polygons created where the two polygon layers intersect (Qiang et al. 2017). As shown in Figure 3, each polygon in the intermediate layer has the attribute information from both layers, including flood zone area (Area_intersect) and which county the polygon belongs to (NAMELSAD20). To calculate the total flood zone area in each county, the dissolve operation can be applied to aggregate the polygons by the attribute county (NAMELSAD20) and sum the flood zone area in each county (SUM_Area_intersect).

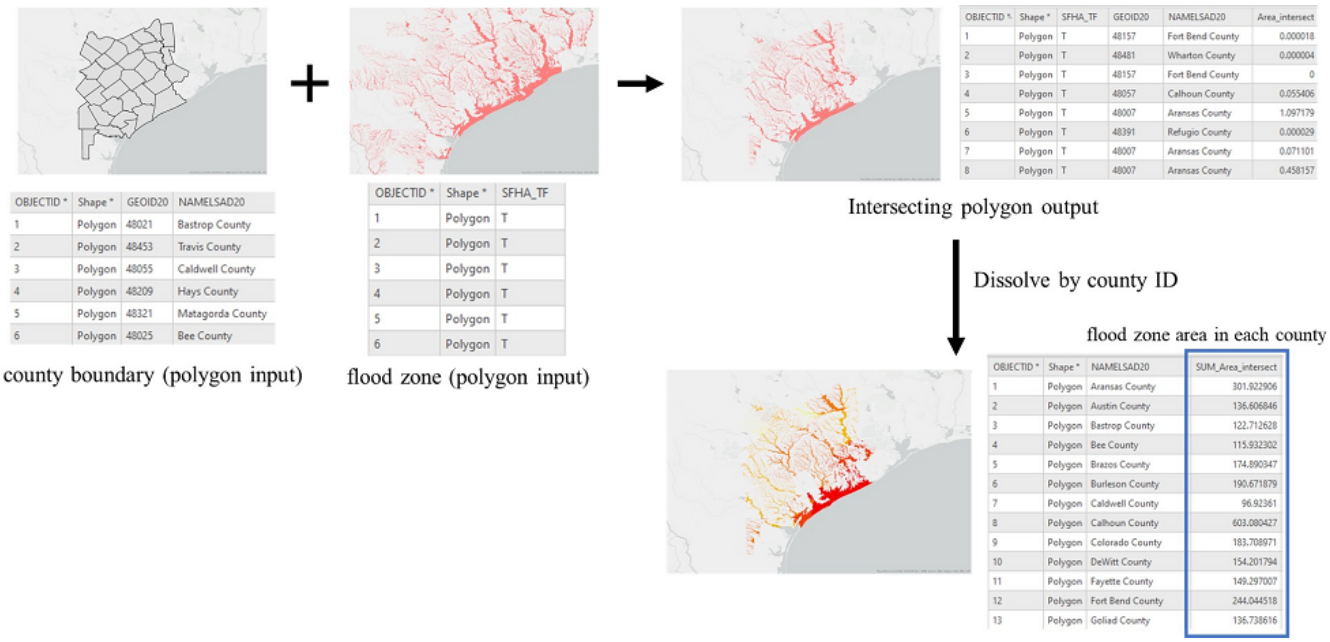



Figure 2. Intersection and dissolve operations using flood zone and county boundaries. Source: author.

In vector overlay, an output layer creates new geometric, attribute, and topological properties. Therefore, this process can result in a substantial increase in output file size when there are large numbers of points, lines, polygons, and attributes that need to be computed in a dataset. Each intermediate layer inherits a combination of geographic and attribute information from input layers. For example, in a line-on-polygon intersection, the line is split by the polygon boundaries. The attribute table of the output line feature layer has the original line attributes and the polygon attributes that the line segment falls within. In this case, vector overlay can become a time-consuming task due to its geometric and computational complexity (Harding et al. 2020).

A common vector overlay error is caused by “sliver polygons”. When the same polygon is presented in different input layers, the boundaries of the same polygon may differ slightly because the data are collected from different sources and processed in different ways.

Therefore, the overlay output will result in numerous small polygons (“sliver polygons”) along the boundary of this polygon after different layers are overlaid (Delafontaine et al 2009). The sliver polygons contain little information but take up significant data size, and dramatically increase processing time. We can reduce the number of sliver polygons by manual editing and automatic removal with a defined snap distance.

### 4. Raster Overlay

Raster overlay combines attributes of two or more raster layers based on map algebra or raster calculus, which computes new raster data through a series of operators (Tomlin 1994). Each cell value represents an attribute value of reality. Compared with vector overlay, raster overlay is relatively simpler and computationally efficient. In practice, vector data can be converted to raster data to reduce the computational burdens. To prevent overlay errors, the input layers must be precisely cell-by-cell aligned and have the same cell size and spatial coverage.

Similar to vector overlay, raster overlay employs a set of basic operators to integrate multiple input layers (Table 3). Boolean operators are used to determine the true or false cells given the criteria AND, OR, NOT, and XOR. Arithmetic operation is also a widely used raster overlay method, which transforms cell values through a variety of mathematical calculations, and thus it cannot be applied to nominal values. Some common operators include addition (+), subtraction (-), division (/), and multiplication (×). The second example in Figure 1 is a raster overlay processing of using the addition operator to identify the point cells falling into an area. Comparison operators are powerful in querying a raster layer based on its attribute values and are always used in tandem with Boolean operators. For example, to answer the first question in Table 3 “Which areas have high population density AND low disaster frequency?”, we can first apply the comparison operator “equal to (==)” in the population density raster layer and disaster frequency layer, respectively. Two intermediate raster layers, one with “population density == high,” and the other with “disaster frequency == low,” can be generated. The Boolean operator AND is then combined with the “equal to” operator to return a binary raster layer in which cells meeting both conditions have true values and the rest have false values. These operators are always combined and work as fundamental tools to implement complex spatial analysis tools, such as zonal statistics, weighted overlay, or weighted sum overlay.

**Table 3. Common Raster Overlay Operations**

| Category | Operators | Application Example |
|---|---|---|
| Boolean | AND, OR, NOT | "Which areas have high population density AND low disaster frequency?" |
| Arithmetic | addition, subtraction, division, multiplication | "What is the total rainfall of each cell in the past 5 years given the yearly data?" |
| Comparison | <, >, =, != | "Where are the areas that changed from land type 'land' to land type 'water' from 2001 (t1) to 2011 (t2)?" |

In practice, the data sources maybe in different formats and we need to overlay raster and vector layers. Depending on the specific tasks, we can either use available tools to overlay raster and vector layers, or convert data from one representation to another one

(rasterization or vectorization). For example, we can use the tool mask to keep the same extent of a raster using a vector layer. In a suitability analysis, we can convert the vector layer of road network into a raster layer and combine it with other raster layers, such as NDVI.

## 5. Overlay Operations in Suitability Analysis

The most common application of overlay operations is suitability analysis. Suitability analysis is an integration of Multi-Criteria Decision Analysis (MCDA) and GIS to evaluate or rank the appropriateness or preference of locations for a specific purpose based on spatial distribution of related characteristics (Malczewski 2014). Some specific applications include habitat suitability assessment for pandas, site selection for a new business, or city expansion planning. GIS enables the spatial integration of data layers, and MCDA provides the theory and methods for analysis design and weight assignment. Each input layer represents a specific factor that is important for the suitability analysis. The output visualizes the suitable or unsuitable areas, in the form of suitability scores or rankings.

The general steps in a suitability analysis are presented in Figure 3 with an example of evaluating suitable places for a new coffee shop. This framework is scalable to suit other applications. There are six major steps in implementing a suitability analysis, including defining research questions, designing decision criteria, preparing the input data, transforming input data, overlay operations, and interpreting the output. Another suitability analysis application using the weighted overlay method to find suitable sites for a new school is illustrated in the GIS&T Body of Knowledge section on Geospatial Analysis and Model Building.

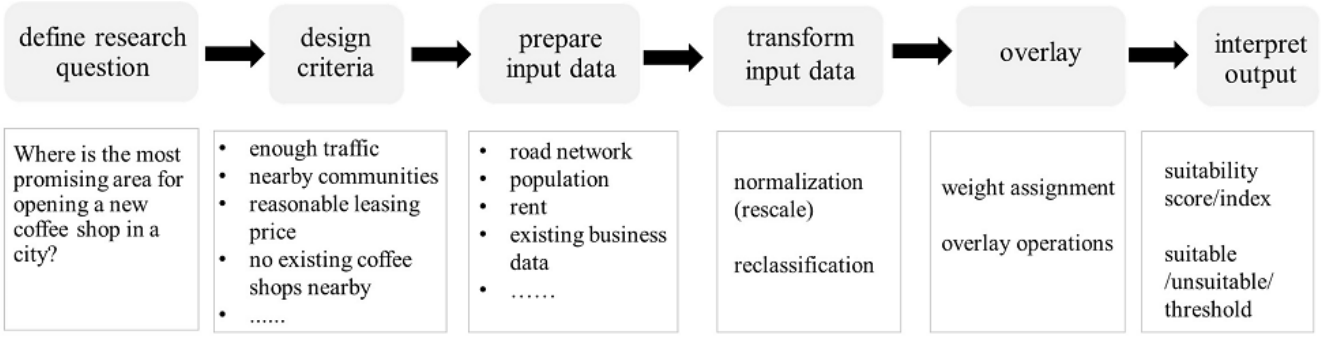



Figure 3. General steps in a suitability analysis. Source: author.

Overlay operations are heavily implemented in suitability analysis studies. Both vector and raster overlay can be used to perform suitability analysis. The decision on which data model to use depends on the research questions, objectives, methods, and data sources. Vector overlay is suitable to conduct suitability analysis that requires clear feature boundaries or accurate distance measurement. For example, when selecting the oil well drilling site in a city, one criterion would be no drilling sites within a certain distance (e.g. 300 meters) of any buildings. This type of criteria requires accurate distance measurement and necessitates vector overlay analysis. Raster overlay is the preferred choice in most cases due to its representation simplicity and processing efficiency. It is used, for example, when

the input data are numeric or categorical factors, the original data are in raster format, or the input layers are complex and large.

### 6. Conclusions

The spatial overlay is a central approach to integrate multiple thematic layers to derive new information and knowledge that can inform multi-criteria decision-making. It plays an important role in many real-world applications, such as disaster vulnerability and resilience assessment, land use planning, business site selection, and resource allocation. This entry outlined the basic principles of overlay operations, the different ways of implementing vector and raster overlay, and the overall workflow of suitability analysis.

## References

Bolstad, P. (2019). GIS Fundamentals: A First Text on Geographic Information Systems, 6th Edition. Acton, MA: XanEdu Publishing Inc.

Delafontaine, M., Nolf, G., Van de Weghe, N., Antrop, M., de Maeyer, P. (2009). Assessment of sliver polygons in geographical vector data. International Journal of Geographical Information Science. 23(6): 719-735.

Harding, T. J., Healey, R. G., Hopkins, S., Dowers, S. (2020). Vector Polygon Overlay. In: Healey, R., Dowers, S., Gittings, B., and Mineter, M. J. (eds). Parallel Processing Algorithms for GIS. CRC Press.

Malczewski, J. (2004). GIS-Based Land Suitability Analysis: A Critical Overview. Progress in Planning 62: 3-65.

Page-Tan, C., Fraser, T., Aldrich, D. P. (2021). Mapping Resilience: GIS Techniques for Disaster Studies. In Rivera, J. D. (ed.), Disaster and Emergency Management Methods. Routledge: 339-354.

Qiang, Y., Lam, N.S.N., Cai, H., Zou, L. (2017). Changes in Exposure to Flood Hazards in the United States. Annals of the American Association of Geographers 107(6):1-19.

Tomlin, C. D. (1994). Map algebra: one perspective. Landscape and Urban Planning, 30 (1-2):3-12.

Unwin, D. (1996). Integration through overlay analysis. In Spatial Analytical Perspectives on GIS. Routledge: 127-138.