# Overlay Analysis

The principle of the overlay of spatial data can be illustrated through the example of producing maps. When printing a map, different layers of spatial information are printed one after the other onto white paper. Each color represents new information on the paper: green represents vegetation, blue for hydrography, and red for the signatures representing residential areas. All the colors together give a very detailed model of reality: the map. This is the main basis of overlays in GIS. First, two or more layers of information from the same area are overlaid onto each other. Then, the topology of the new layer is updated: if a point now lies within a polygon, it gets assigned this information as a new attribute. If two lines intersect ("arcs") a new node will be added at their intersection. If two polygons intersect, a unique identification number is given to the intersecting set and so forth. Ultimately, the overlay results in an information gain. In order for this integration to make sense, all input layers must have the same reference system and scale. The map will only be legible if all the layers fit together exactly in regards to position and scale. The process itself is independent of whether a raster or a vector model is used. With a raster model, the overlay operation is rather an overlay than an intersection. The integration of information from various sources through overlay is one of the most important functions of a GIS.
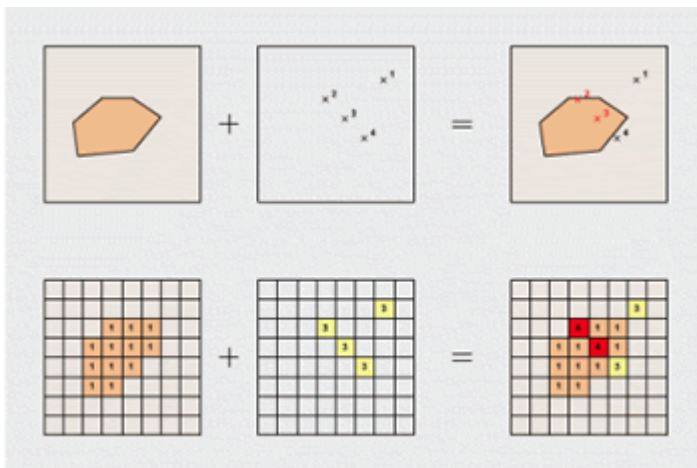
Point-in-polygon: in the vector model, this overlay determines the points lying inside a specific polygon. In the example, we are looking for all hotels that are located in the settlement areas. In the resulting layer, the points have the additional information whether or not they are in the settlement area. In the raster model, the points in question are visible through the addition of the two input layers.
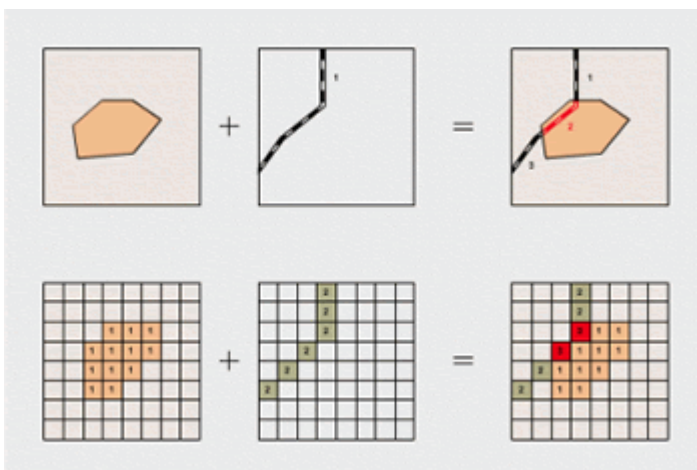
Line-in-polygon: the overlay of lines and polygons is more complex. The example shows the calculation of road sections located in the settlement area. In the vector model, the topology changes: the original line is cut into shorter segments by the intersection points. It has to be specified for each segment whether it is inside or outside the settlement polygon. In the raster
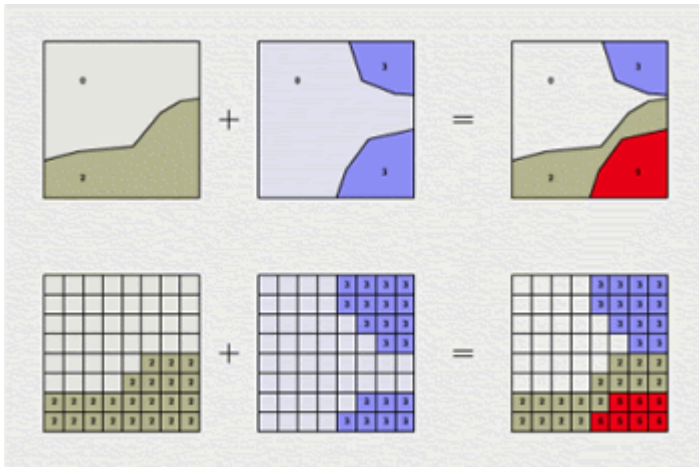
model, a simple addition identifies the interest areas.

Polygon-on-polygon: in the vector model, the most complex case is the intersection of polygons. The result is a data layer with a whole new topology. The overlay of contour lines results in a variety of new intersections and polygons for which the attributes have to be reassigned. In addition, for non-cut areas it has to be checked whether they contain areas of another information layer i.e. whether island polygons were created. Raster overlay is very simple, however. Again, the cell values of the input layers are calculated.

Overlays, and in particular Boolean overlays, are possible using both vector and raster models. In general, vector overlay is more complex, time-consuming, and computationally intensive than raster overlay which is easier, faster, and more efficient. Nevertheless, before a raster overlay can be performed, a set of elaborate preprocessing steps must be conducted to approximate the cost for the two methods. There are other reasons to opt for one or the other of the two data models. The following list shows when each model should be considered.

Vector Overlays:

The first type of overly operations we're going to discuss in this section are vector overlay. A vector overlay involves combining point, line, or polygon geometry and their associated attributes. All overly operations create new geometry and a new output geospatial data set. You should be cautioned that with certain overly operations, very large attribute tables may result if the overly operations combine many layers, and each layer has a very large attribute table. Additionally, it might be possible that the combined attribute tables would cause duplicate attribute fields to exist. In these cases, you should consider reducing the number of transferred attributes to the minimum required, and renaming duplicate fields so that there is no ambiguity.

1. The first basic case of overlays is the clip function. The clip function defines the area for which features will be output based on a "clipping" polygon. I like to think of it in terms of cookie dough, and a cookie-cutter. The cookie dough is the data layer that we want to reduce in size based on our cookie-cutter. The clipping layer is the cookie-cutter that will determine how we cut out the part of the cookie dough we are interested in. It is important to note that in the clip operation, only the geometry and attributes of the data layer are transferred to the result layer. Neither the geometry or

attribute of the clipping layer are included in the output, just as you would not make your cookies in the oven with the cookie-cutter still surrounding the cookie dough.

2. The next basic case of overlay is the intersection operation. The intersection operation combines data from two or more input layers, but the combination only exists for the regions where all input layers contain data. This is similar to the clip function, however the attributes and geometry of all input data layers are transferred to the result layer.

3. The third and final basic case of overlay is the union overlay operation. The union operation is an overlay that includes all data from all of the input data layers. Since the union operation combines all attributes, and all geometry, no geographic data are discarded in the union operation, and all corresponding attribute data are saved for all regions.

When performing vector overlays, there are a few problems that you should be aware of. In the case where the input features represent common features that are represented in both layers, they may have slightly different geometry. This creates sliver polygons when the overly operations are performed. For instance, imagine we have one input data set the represents the national boundary of Canada created at a 1 to 5,000,000 scale. And we have a second input data set of the national boundary of the United States of America created at 12 1 million scale. If we stack these two layers on top of each other, it is probable that the boundary lines will not overlay exactly, as they were created at different scales, and therefore the Canadian boundary line has probably been simplified more than the United States boundary line. The areas where the two input data sets do not match, or result in very small, typically insignificant sliver polygons that need to be dealt with. There are several methods that exist to reduce the occurrence of sliver polygons, such as manually snapping the boundaries together, or introducing a fuzziness to the edges of the overly operation, or simply manually deleting the sliver polygons.

Advantages of vector overlay:

- Clear, sharp-edged criteria, e.g. no-building zones in a hazard map where you are not allowed to build under any circumstances;
- Clear, sharp-edged input data;
- Clear distance criteria, e.g. maintaining a minimum distance, "at least 100 meters away from the settlement area."

Advantages of raster overlay:

- Fluffy or fuzzy criteria, e.g. the criterion "steep" is difficult to clearly delineate;
- Continuous and arguable data;
- The creation of cost surfaces.