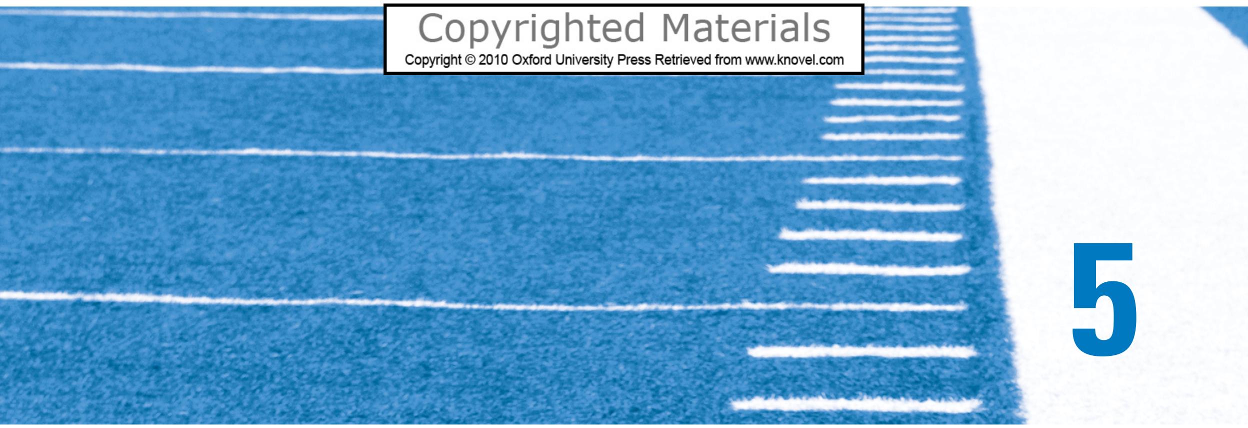

5

Combining data layers

5.1 Introduction

This chapter is concerned with the nature of connections between objects in terms of overlap or proximity between different objects. Operators that deal with overlaps between objects require information on where specific features overlap in space—the overlap of lines is termed the 'intersection'. Such operators provide the basis of multi-criteria decision analysis (MCDA, see Section 5.3) whereby information in multiple data layers is taken into account. Overlay operators are a key component of GIS software packages and they enable the identification of features that share the same geographical space (whether points, lines, or areas). The main focus in this chapter is on vector data; Section 10.2 deals briefly with raster overlay.

Exploring connections between objects may provide the basis of many different kinds of analyses. At a simple level, for example, there may be a need to identify which areas border another area (i.e. which areas are contiguous). Using topological structures, as detailed in Section 2.2.3, this kind of information may be obtained directly. Often the concern will be to identify which objects contain other objects or are contained by other objects (e.g. which local government boundary contains which houses). In many cases, connectivity between different kinds of objects is of interest. Finding the common areas from two sets of polygon features (e.g. highly populated areas and highly polluted areas) is one kind of application that is frequently encountered in GIS contexts. This chapter is concerned with overlaps between features in different data layers.

5.2 Multiple features: overlays

Overlay operators combine information in various ways from two or more sets of spatial data. Such operators are concerned with inclusion and with overlap or intersection (Burrough and McDonnell, 1998). For properties (e.g. areas) A and B, the two can be defined as follows:

Inclusion Is A contained within B? Operators transfer attributes to the features contained or that contain other features. For example, if point A is contained within polygon B, then the point is labelled as belonging to area B and attribute information from B is transferred to A.

Intersection Do A and B overlap? This leads to the creation of new spatial features. For example, if two areas overlap then at the points where the edges overlap new nodes are created and overlapping areas in the output contain the attributes of both input layers (in this example, they have type A AND type B). Inclusion and intersection are both illustrated below.

Overlay operators can be used for polygon overlay, line-in-polygon overlay, point-in-polygon overlay, and to identify overlapping lines. Polygon overlay is a spatial operation that overlays one polygon layer on another to create a new polygon layer. The spatial features of each set of polygons (or a subset) and their polygon attributes are joined in the output layer. Joining polygons enables the use of operations requiring new polygon combinations (e.g. all areas that are both highly populated and highly polluted). Line-in-polygon operations allow the line features to take the attributes of the polygon in which they lie (e.g. identification of which census areas a road passes through). Point-in-polygon overlay transfers the attributes of the polygon in which the point lies to the point (e.g. labelling a house with the administrative region it is contained by). These operations (including point-in-polygon overlay) necessitate identification of line intersections. A means of identifying line intersections is outlined in Appendix D.

5.2.1 Point in polygon

Ascertaining which polygon a point falls in is a frequent problem in GIS contexts and questions such as 'Which disease events are in which administrative areas?' are often posed. There are various ways in which this can be resolved. Laurini and Thompson (1992) outline one approach. The essence of this algorithm is that a line (the 'half line') is drawn from the point to the edge of the map and the number of intersections with lines that are part of the current polygon is counted. The number of intersections will only be odd if the point falls within the boundary of the current polygon. This procedure is illustrated in Figure 5.1. In this example the half line crosses the boundary of the polygon in which it sits three times, confirming that is where it is located. The search process can be accelerated through the use of a minimum enclosing rectangle (MER, as defined in Appendix D) to ascertain if the point can possibly be contained within a given polygon (i.e. if the point is within the MER then it may also be within the polygon contained by the MER). Once polygons are excluded in this way, the remaining candidates can be tested using the line segments (see Appendix D for a testing procedure) until the relevant polygon is identified. Heywood *et al.* (2006) detail problem cases, for example where the point lies on a polygon boundary or the half line

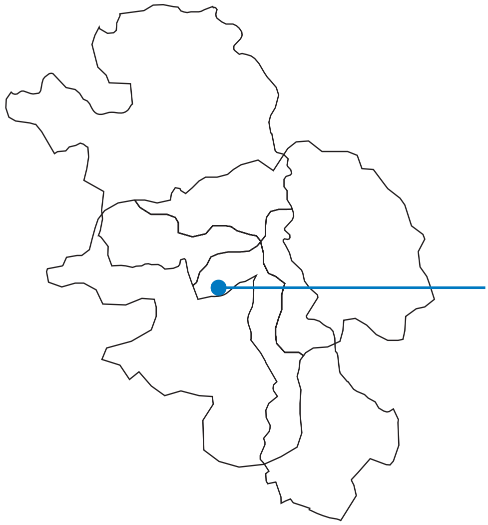

Figure 5.1 Point-in-polygon test example.

coincides with a boundary. Such cases are dealt with through additional stages added to the algorithm employed.

The point-in-polygon operation can be extended to the idea of assessing if an island polygon is contained within another polygon. Firstly a MER can be used to assess if containment of the island polygon by another polygon is possible. Next, a series of horizontal bands is drawn across the polygon that may contain the island polygon. The first and last locations on these lines that intersect the island polygon are then treated as for a point (as shown above). If the number of intersections for each line is odd then the polygon is enclosed completely (Burrough and McDonnell, 1998).

5.2.2 Overlay operators

This section deals with methods that enable answering of questions such as 'How many sites of special scientific interest (SSSI) occur in areas with highly polluted soil?'. This kind of question is well depicted using Boolean logic (see Section 2.11.1). If SSSIs are termed 'A' and highly polluted soil is termed 'B' then the statement A AND B (also given by $A \cap B$) would correspond to our question: all areas fulfilling both criteria are selected. In Appendix D, an approach to the line intersection problem is outlined and this approach can be employed to implement the methods detailed in this section.

Polygon overlay operations are defined and illustrated below. These approaches assume that the features of interest (e.g. polygons representing SSSIs and those representing areas with highly polluted soil) are stored in separate layers. In turn the Union, Intersect, and Identity operations are outlined. These operations are similar, differing only in the spatial features that remain in the output data layer. In all cases, an input layer and a second overlay layer are shown along with the overlay output.

Union overlay joins two sets of polygons and retains all areas from both layers (and the attributes in both layers are kept), so it makes no practical difference which is the

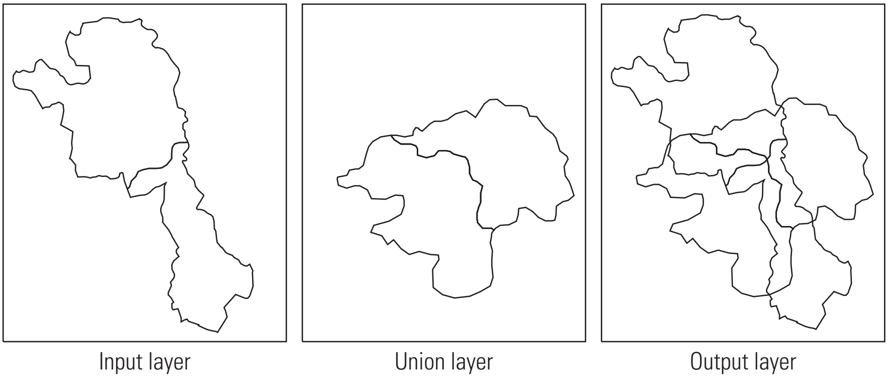

Figure 5.2 Union operator.

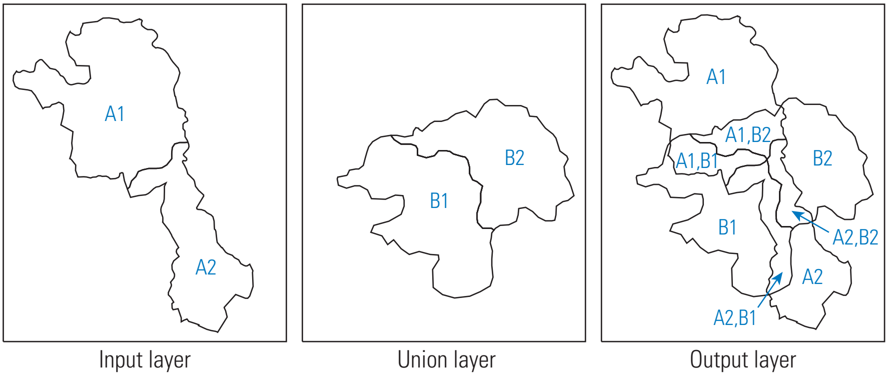

Figure 5.3 Union operator: transfer of attributes.

input layer and which is the union layer. The operation corresponds to the Boolean OR and is illustrated in Figure 5.2.

The attributes from the two layers are joined as shown in Figure 5.3. In that case, the first layer has polygons that each represent one attribute labelled A1 or A2. Similarly, the polygons in the second layer represent attributes labelled B1 and B2. In the output (union) layer those overlapping areas take both sets of attributes. In this way, new polygons with particular combinations of attributes are generated. Of course, each input polygon may contain many attributes, which will be transferred in the same way.

Intersect overlays points, lines, or polygons on polygons but retains only those portions of the input layer falling within the overlay (intersect) layer features. The intersect overlay corresponds to Boolean AND (for layers A and B, given by $A \cap B$) and is illustrated in Figure 5.4.

For lines, line segments in the input layer that fall within the intersect layer are retained. For points, those points in the input layer that are located within the intersect

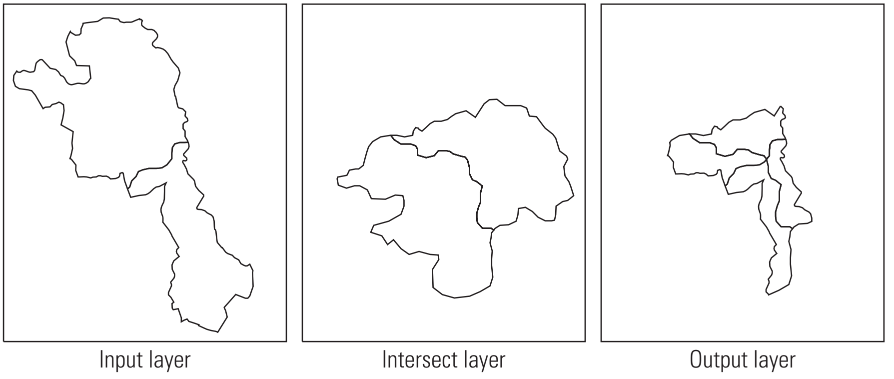

Figure 5.4 Intersect operator.

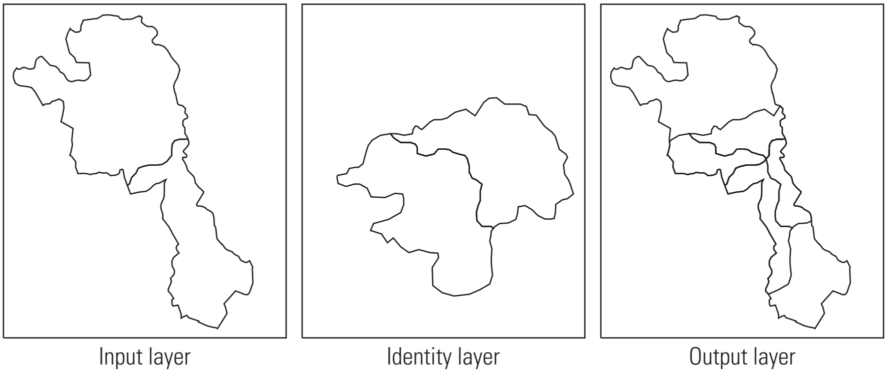

Figure 5.5 Identity operator.

layer features are retained. Attributes from both layers are transferred to the output layer.

Identity overlays points, lines, or polygons on polygons and keeps all input layer features (with transfer of attributes for both layers). The identity overlay is illustrated in Figure 5.5.

For lines, all input lines are retained but the lines are split and nodes added where the lines overlap with the identity layer features. For points, all input points are retained but the ID numbers of the polygon/s within which points fall are assigned to those points.

5.2.3 'Cookie cutter' operations: erase and clip

'Cookie cutter' operations are widely used—these enable cutting features within a given feature (termed the 'erase operator') or outside a given feature (sometimes termed the 'clip operator'). Unlike the operators described above, attributes are not transferred from both input layers to the output layer—instead only the attributes of the input layer appear in the output layer.

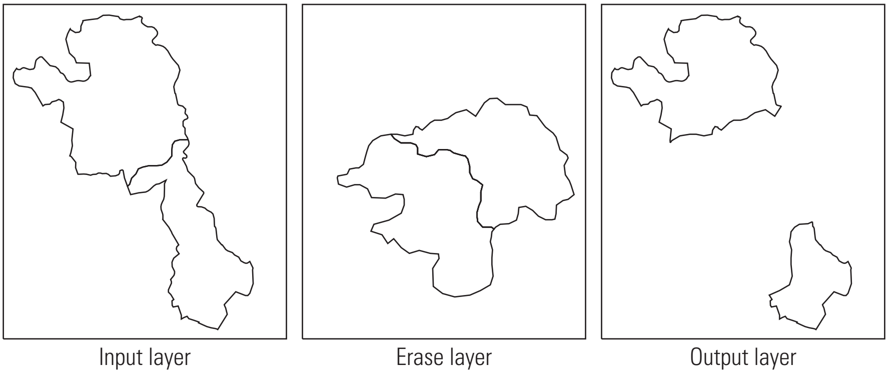

Figure 5.6 Polygon erase.

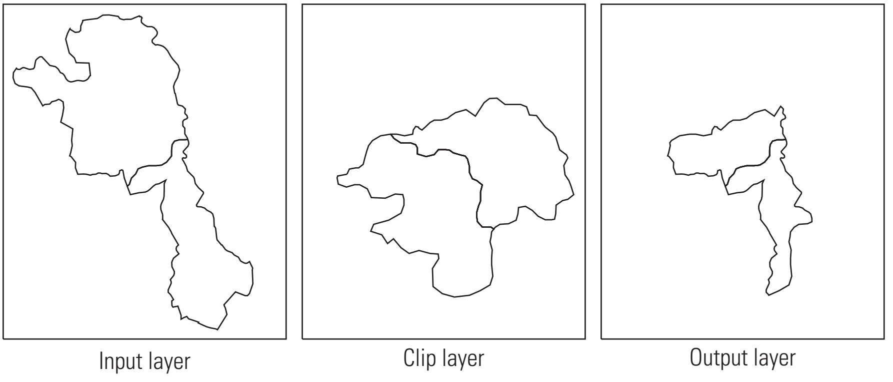

Figure 5.7 Polygon clip.

Erase creates a new layer by overlaying two sets of features. The polygons of the erase layer define the erasing region. Input layer features that are within the erasing region are removed—this is illustrated in Figure 5.6. The erase operator can be used with polygons, lines, or points as inputs.

The clip operator is similar to erase except that the features that are within the clip region are preserved. Like the erase operator, the clip operator can be used with polygons, lines, or points as inputs. The clip operator is illustrated in Figure 5.7.

5.2.4 Applications and problems

Overlays of various kinds are central to many GIS-based projects. A large proportion of applications that utilize multiple spatial attributes make use of overlay procedures. As an example, Sprague *et al.* (2007) are concerned with assessment of the persistence of rice paddies within the Kanto Plain of Japan. In that study, historic (i.e. late nineteenth century) maps were georeferenced to a modern map using features that appear on both the historic maps and modern maps. Overlay procedures were then used to combine rice paddies represented on the historic and the modern maps. It was then possible to identify the common areas of rice paddies on the two maps and to ascertain which areas no longer contain rice paddies.

Combining different data layers may be problematic where polygon boundaries in the input layers are supposed to be identical but in fact differ. The end result is termed 'spurious (or slither) polygons'—(usually) small polygons that represent the difference between the two sets of boundaries. Various approaches exist for removing such spurious polygons. If there is greater confidence in the accuracy of one set of boundaries than the other then the boundaries with greater accuracy may be retained in preference to the other less accurate boundaries.

5.3 Multicriteria decision analysis

GIS allow the integration of diverse data sources and facilitate the exploration of the relationships between them. At a simple level, it may be desired to identify a set of areas that fulfil several dozen criteria (e.g. areas more than a given distance from some feature, with gradual slopes, a particular soil type, and so on). Earlier sections introduced tools, such as buffers (Section 4.5) and overlay operators (previous section), which can be used to help address such issues. This section expands on some key topics that relate to considering multiple criteria in this way. The term 'GIS-based multicriteria decision analysis' (GIS-MCDA) encapsulates the processes involved in using GIS to help decision making using multiple data sources. Malczewski (2006) provides a simple definition of GIS-MCDA as 'a process that transforms and combines geographical data and value judgements (the decision-maker's preferences) to obtain information for decision making' (p. 703). With GIS-MCDA, the relative importance of different criteria can be taken into account. For example, if, in some planning process, accessibility by road is more important than the slope of the terrain then proximity to roads may be given a larger weight than slope as a criterion in the decision-making process.

A means of selecting particular alternatives from a set of available options is a decision rule (Malczewski, 2006). The weighted summation decision rule approach and similar approaches are the most commonly applied in the literature (Malczewski, 2006). Malczewski (1999) outlines one simple additive weighting method. Using this approach each criterion layer is standardized (e.g. by dividing each value in a layer by its maximum value). This is necessary to enable comparison of like with like as direct comparison of, say, distance from roads with slope of terrain would be meaningless—standardization means that all units will be comparable (using the approach suggested above they will all range from 0 to 1; see Heywood *et al.* (2006), pp. 239–240, for another example). Next, weights are determined (this may be quite a subjective procedure). If a layer is to be assigned 40% of the weight this can be expressed as 0.4, with the weights for the other layers totalling 0.6. Then, each of the standardized map layers is multiplied by the weights and the weighted standardized maps are added together. The optimal alternative (e.g. most suitable area or areas for development) is

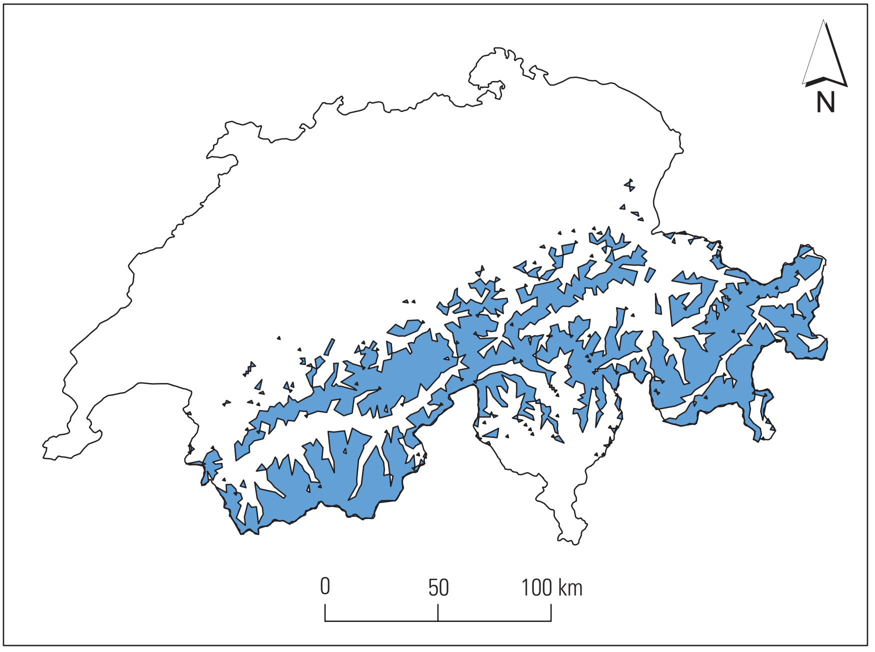

Figure 5.8 Areas in Switzerland with elevation values greater than 2000 m.

that which has the largest score (and thus the highest rank). All that is needed to implement such a system are simple overlay tools and such an approach can readily be applied using vector or raster data (Malczewski, 1999), although most MCDA applications are conducted in the raster environment (see Section 10.2 for a related discussion).

5.4 Case study

This case study makes use of zones (polygons) representing areas in Switzerland with (1) elevation above a threshold value and (2) predicted precipitation amount above a threshold value. Applications of this kind are common in GIS contexts and the results of the analysis could easily be combined with data on other properties such as protected areas or administrative zones. The data on which this analysis is based are described in Section 8.7. Figure 5.8 shows areas with elevations of greater than 2000 m while Figure 5.9 shows areas with a predicted (daily) precipitation amount of greater than 250 1/10 mm (i.e. 250 tenths of a millimetre). In both cases, the vector features were generated from raster grids using the ArcGIS™ software and the edges of the polygon outputs were generalized (smoothed). The aim was to find the common areas and for this purpose the intersect overlay operator was used. The final result (areas that fulfil both criteria) is shown in Figure 5.10.

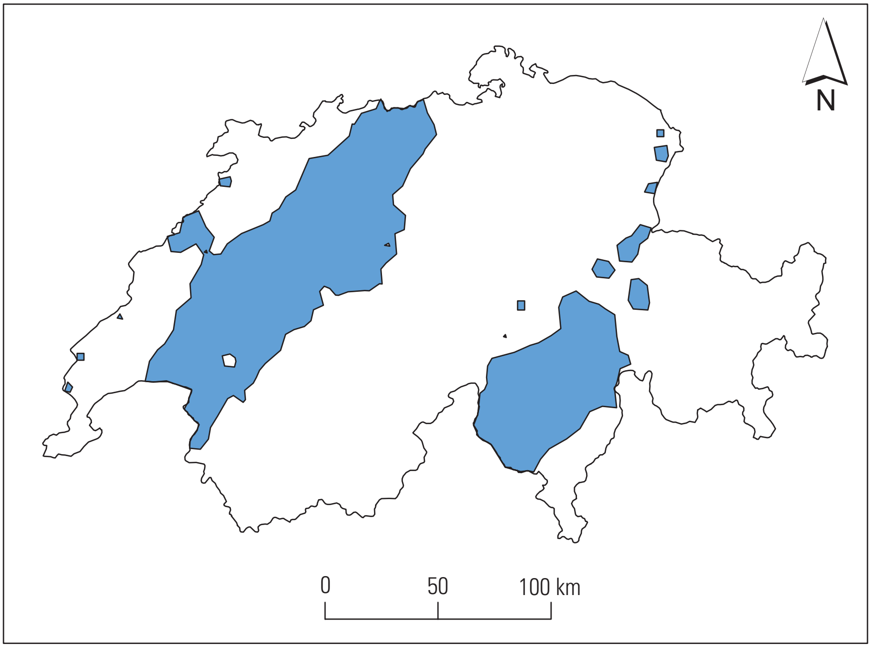

Figure 5.9 Areas in Switzerland with predicted daily precipitation amount values greater than 250 1/10 mm.

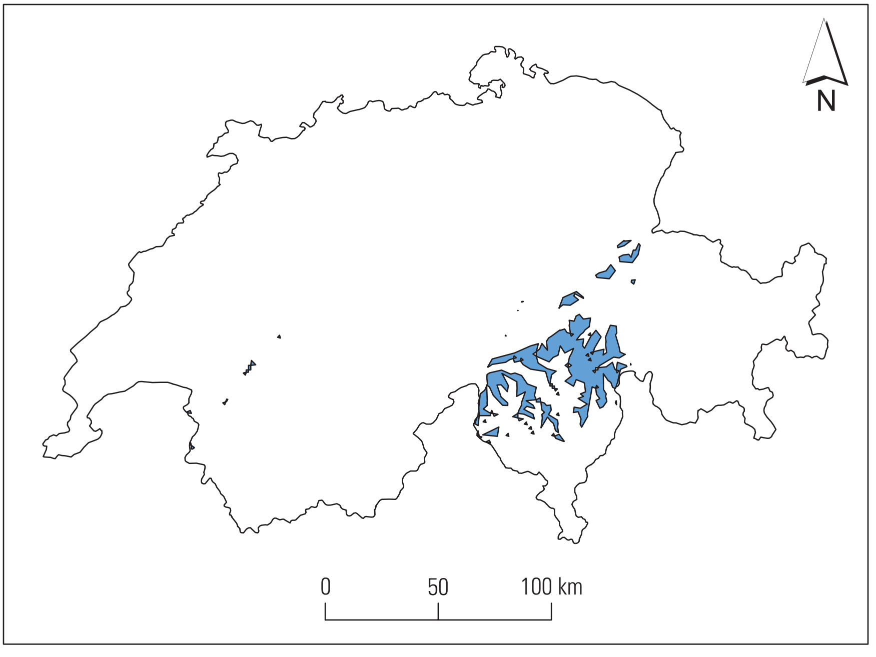

Figure 5.10 Areas in Switzerland with elevation values greater than 2000 m AND daily precipitation amount values of greater than 250 1/10 mm.

Different methods for prediction of precipitation amounts will result in potentially very different outputs (see Chapter 9). Figure 5.10 includes some very small polygons that may not have existed if a different interpolation procedure was used (or the inputs had varied in some other way) and, in such cases, removal of these small polygons might be considered.

Summary

One of the most widely exploited benefits of GIS is their capacity to combine multiple data layers in flexible ways. Such functionality allows complex multiple criteria to be taken into account simultaneously in a way that was practically impossible before the advent of computer-based systems for spatial data analysis. This chapter has provided an overview of some key ways of identifying overlaps between features in different data layers and for combining data layers. In addition, the identification of areas which fulfil multiple criteria was discussed.

Further reading

Overlay operators are described in standard GIS textbooks such as those by **Burrough and McDonnell (1998)**, **Heywood *et al.* (2006), Chang (2008)**, and **Longley *et al.* (2005a)**. **Chou (1997)**, **Lee and Wong (2000)**, and **O'Sullivan and Unwin (2002)** describe the principles in more detail. **Wise (2002)** details some key algorithms for overlay. A key reference for GIS-based multicriteria decision analysis is the book by **Malczewski (1999)**.

➜ The next chapter is concerned with network analysis and introduces tools to address questions like 'What is the shortest route between two places on a road network?'.