Subject	Geology
Paper No and Title	**Remote Sensing and GIS**
Module No and Title	GIS Spatial Overlay Analysis
Module Tag	RS & GIS XXXI

Principal Investigator	Co-Principal Investigator	Co-Principal Investigator
Prof. Talat Ahmad ***Vice-Chancellor*** Jamia Millia Islamia Delhi	**Prof. Devesh K Sinha** Department of Geology University of Delhi Delhi	**Prof. P. P. Chakraborty** Department of Geology University of Delhi Delhi
Paper Coordinator	**Content Writer**	**Reviewer**
Dr. Atiqur Rahman Department of Geography Jamia Millia Islamia Delhi	**Dr. Manika Gupta** Department of Geology University of Delhi Delhi	**Dr. Atiqur Rahman** Department of Geography Jamia Millia Islamia Delhi

1. Learning Outcomes

After studying this module you shall be able to:

- Understand the role of overlay methods in GIS for spatial decision making
- Learn about different overlay methods like boolean and weighted approaches
- Understand different overlay operations to combine two or more data layers

2. Introduction

GIS, geographic information system is widely utilized for spatial analysis and decision making like flood vulnerability mapping, water resource management etc. Spatial overlay analysis in GIS is one of the fundamental operations that can be performed on both Raster and vector data. Spatial overlay analysis as the name suggests, is utilized to combine the features of two or more layers within the same geographical boundaries and provide a changed entity as required. Spatial overlaying is independent of model being vector or raster. However, to overlay to or more layers, it is important to note that all the layers ought to have been registered and have same coordinate system. This technique is also useful for delineating hazard zones, vulnerability mapping and even groundwater potential zone.

3. Vector and raster overlay

Overlay method is distinguished based on the feature of one the layers, i.e. whether the feature is point, line or polygon. Thus, there can be three different overlay methods which are given as below

3.1 Point-in-polygon

The point-in-polygon overlay determines the points from one input layer lying inside a particular polygon. In this method the output layer is still a point layer but has added attributes of the polygon feature from the second layer. The Fig. 1 shows an example of point-in-polygon, where in point layer shows soil type attribute represented by points, 1,2 and 3 and second layer is polygon layer, which shows three types of vegetation, namely A, B and C. So, if we want to also know the vegetation type associated with the soil type, feature overlay is utilized. Thus, here the output layer is again point layer but now the attribute information has additional data representing the vegetation for each point determined by overlaying the two layers. Therefore, the attribute information is changed from 1,2 and 3 to 1A, 2B and 3C respectively.

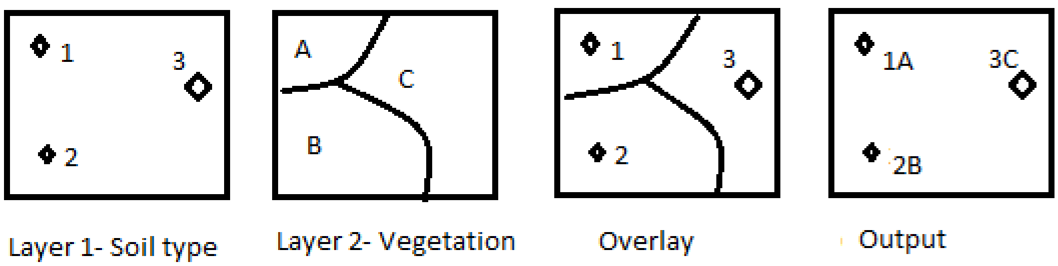

Fig. 1 Point-in-polygon feature overlay

3.2 Line-in-polygon

This method is similar to the previous one, only difference being that the one input and output layer is now line layer. Here the line feature can be divided into multiple segment depending the polygon it lies within. As shown in Fig. 2, the input layer represents road feature and second layer may represent urban settlement. Now, overlay of the two layers may help in determining, which section of the road is within urban settlement polygon and which is outside the polygon. Thus, the output layer is again a line layer but the one segment is now divided in three segments with additional attribute information of settlement.

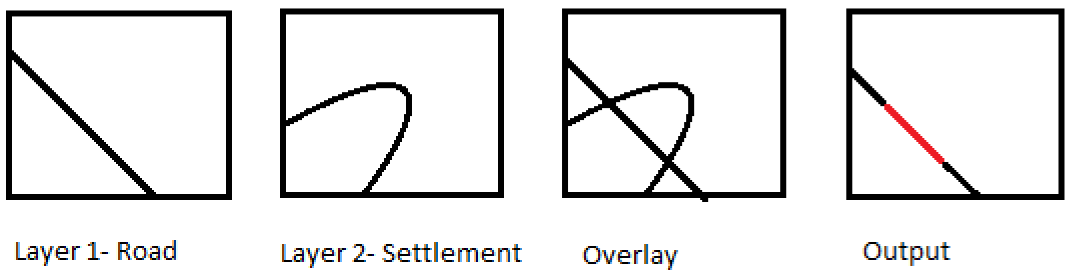

Fig. 2 Line-in-polygon feature overlay

3.3 Polygon-in-polygon

This feature overlay allows the overlaying of two polygon dataset. Thus, the new output layer has the added information of both polygon features. A very simple example has been illustrated in the Fig.3 wherein the polygon are represented by straight lines. However, in real situation examples, the polygons may not be straight and increase the complexity of the computational analysis. In such cases, a variety of new intersections are obtained and new polygons will be formed and attribute data is reassigned. In this overlaying example, Layer 1 may represent suppose soil classes like clay soil, sandy soil etc. and Layer 2 may represent maybe land cover like agriculture or forest. Thus, the output layer has attributes added for both the layers and we can exactly know the both soil type and land cover from a single thematic map.

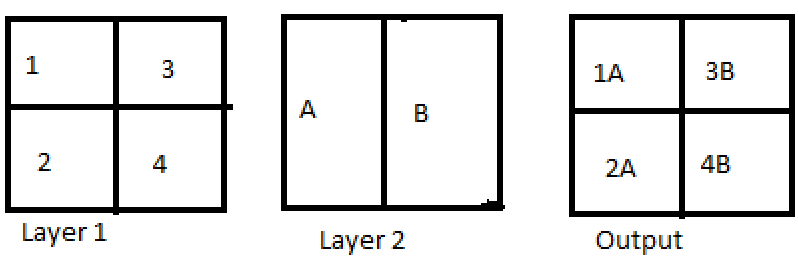

Fig. 3 Polygon-in-polygon feature overlay

4. Overlay methods

The overlaying methods involve operations like Boolean algebra comprising the logical comparison between the layers. The Boolean algebra, basically takes back to Venn diagrams, which were utilized to identify "union", "intersection" and "difference" features within two layers. The Boolean algebra method allows the decision making for the categorical data. This overlaying method allows creation of conditional statements, which results in a true/false output layer. It will be seen that we can achieve different results from the same two layers based on the Boolean method employed. The various Boolean methods employed for spatial overlaying is explained below-

4.1 Boolean methods

4.1.1 Union (OR logical)

Union overlay implies the preservation of features of both the concerned layers. Thus, the new layer would have information of both the layers and is independent of the covered area. Union is represented by the "OR" criteria. As shown in Fig. 4a, if we have two layers A and B, the output layer is represented by A+B. As an example, if we have a road layer, A and land use layer, B then the overlay of layers results in layer, C, shown in Fig 4b. The union Boolean method can also be utilized to obtain output layer such that it represents all areas that meet either the first or the second criterion or both that is at least one criterion has to be true, like if we want to check agricultural site or clay soil type, shown in Fig. 4c.

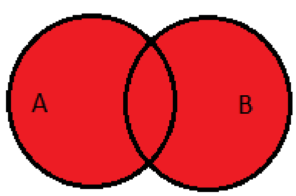

Fig. 4a Set Union, True condition represented by red color

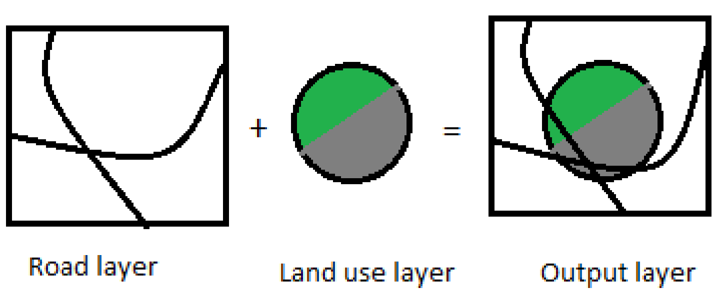

Fig 4b. Union of two themes

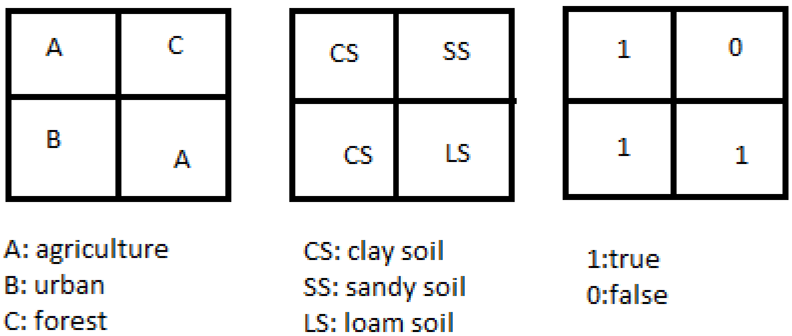

Fig 4c. Output layer showing union(1), for either agriculture or clay soil or both

4.1.2 Intersect (AND logical)

Intersection overlay is utilized, when the information to be preserved falls within the required categorical data. The intersect option is based on the AND Boolean criteria, which implies both first and second criterion needs to be met. If represented by a Venn diagram, the criterion points to the shared part of set A and set B, shown in Fig.5a. Intersection may be important, if for example we need the clay soil type only within the agriculture class. So, here the output layer would only have true for the cell, which satisfies both agriculture class and clay soil type, as shown in Fig 5b. Thus, the true option can be seen only for one cell as compared to true for three cells in Fig 4c, highlighting the difference between union and intersection. The criteria can also be utilized to find the erosion potential of the site under consideration. Here, if layer 1 represents slope and layer two presence of vegetation. The criteria for erosion would be slope greater than 10 degree in layer 1 and no vegetation in layer 2.

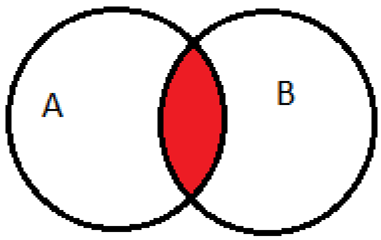

Fig 5a. Set intersection, True condition represented by red color

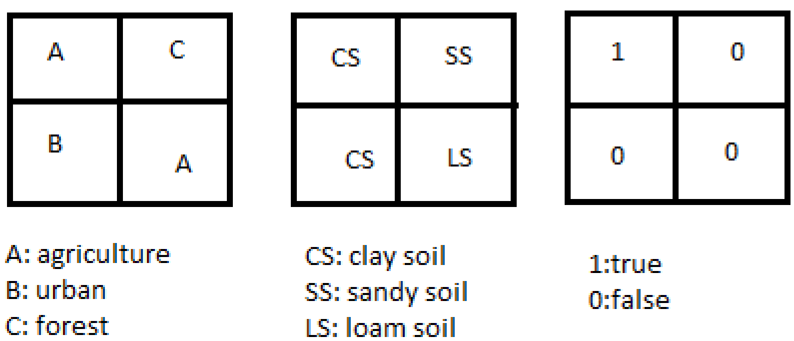

Fig 5b. Output layer showing intersection(1), for both agriculture and clay soil

4.1.3 Exclusive disjunction (XOR logical)

Exclusive disjunction is represented by XOR criteria. The analysis creates a layer that represents area, which meet either of the criteria but not both. The Boolean method is represented in Fig 6a. through a Venn diagram, where the output layer is shown with either A or B being true but not both. So this implies, the intersection is excluded. Again considering the previous example of two layers, land use layer and soil layer, we can identify, which areas are either agriculture or clay soil but not both, as shown in Fig 6b. It can be seen that the output is now true for two cells that is either agriculture area or clay soil area but not both.

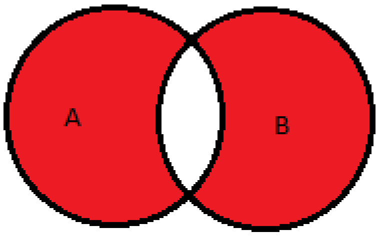

Fig 6a. Set exclusive disjunction, True condition represented by red color

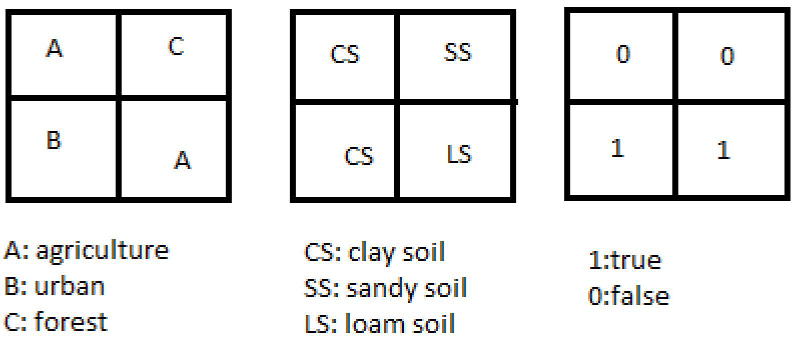

Fig 6b. Output layer showing exclusive disjunction(1), for either agriculture or clay soil but not both

4.1.4 Negation (NOT logical)

Negation or NOT Boolean method, represents the portion excluded from the set A as shown in Fig. 7a. In considering two or more spatial layers, criteria implies we need the first criteria to be met with second criteria being negated. Continuing with the two layer example from above section, we employ negation such as agriculture is true but soil layer should not be clay soil. The output of this condition is given in Fig 7b. The negation method illustrates the utilization of absence of certain criteria on the output layer.

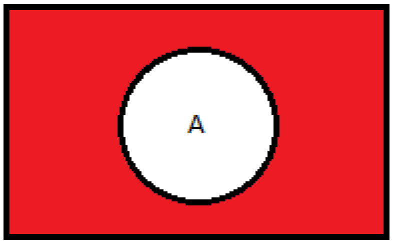

Fig 7a. Set negation, True condition represented by red color

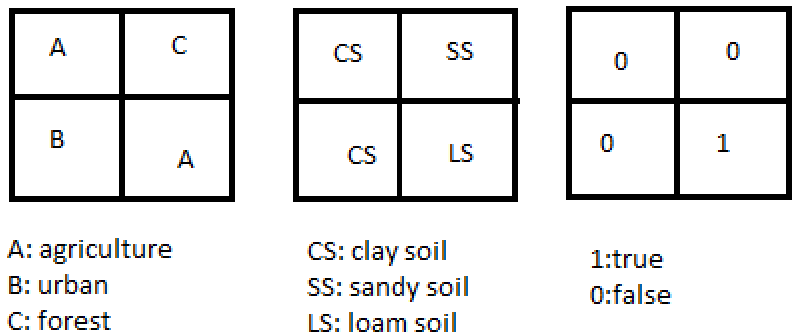

Fig 7b. Output layer showing negation(1), for agriculture true and clay soil false

4.2 Weighted overlay methods

The Boolean method as seen allows the criterion considered for each layer to have equal importance. For example, while considering erosion potential, we might want to give more importance to steepness of the slope as compared to the vegetation. The principle of not considering each input layer equally and assigning weights according to relative influence level of each layer is known as weighted overlay. Weighted linear sum is one of the approaches that is applied to a raster dataset, grid cells are multiplied by the weights assigned to particular layer and then the summation is done. The weighting can be done according to the methods given below -

4.2.1 Weighting by ranking

Weights by rank allows to provide a rank to each of the input layer either in ascending or descending order of significance of the given layer. The weightage now can be estimated either x layers as

I. if rank is y than the weight is x-y+1
II. if rank is y than the weight is 1/y
III. if rank is y than the weight is $(x-y+1)^n$, where n determines the weights.

Even though the method is easy to implement, the relative importance between layers start decreasing with the increase in number of layers.

4.2.2 Weighting by rating

Again in this approach the layers are first ranked and then each layer is given a score. The overall score is distributed within the chosen criteria. For example, with a total score of 100, criteria could receive the weights 40, 25, 20, 10, and 5 points with decreasing importance or the most important criterion is given a score of 100, while the insignificant criteria is given 0. However, this method is highly subjective.

4.2.3 Multi criteria evaluation (MCE)

The third criteria of weighting is known as Multi criteria evaluation and is based on the Analytic Hierarchy Process (AHP), given by Saaty (1980). The method utilizes a pairwise comparison procedure for the derivation of factor weights. Using a 9-point rating scale, each layer is compared to each other for its relative importance in developing the final solution. The principal eigenvector of this matrix is then calculated to derive the best-fit set of weights. To explain the method, the steps followed in the method are given below

1. The suitability of factors are set to a common criteria so as to make the comparisons possible. Thus, for example, if we have two layers slope (%)

and elevation (m), the values are converted in a common range 0-255, such that 0 is less suitable while 255 is highly suitable.

2. A matrix is constructed, where each criterion is compared with the other criteria, relative to its importance, on a scale from 1 to 9. Here, 1 = equal preference between two factors; 9 = a particular factor is extremely favored over the other. The weight estimate is calculated by obtaining the eigen values. Here, first the factors are normalized by dividing each cell by the column total and then the mean value is obtained for each of the rows representing the factors.

   Table 1 Comparison matrix

	C1	C2	C3
C1	1	9	3
C2	1/9	1	1/5
C3	1/3	5	1
Total	1.444	15.000	4.200

   Table 2 Weight determination

	C1	C2	C3	Priority vector
C1	0.692	0.600	0.714	0.669
C2	0.0769	0.067	0.048	0.063
C3	0.231	0.333	0.238	0.267

Once the weights are determined, the consistency ratio (CR) is calculated as, CR= Consistency Index (CI)/Random consistency index (RI)

where in,

$$CI = (\lambda_{max} - n)/(n-1)$$

λ_{max} is principle eigen value, calculated as summation of product of column total from table 1 and priority vector in table 2. If CR > 0.10, then some pairwise values need to be reconsidered and the process is repeated until the desired value of CR < 0.10 is reached.

3. Finally the criteria is aggregated

$$x = \sum w_i n_i \ \mathrm{x} \ \prod \mathrm{c_j}$$

where,

x is final suitability score

w_i is weights of each factor

n_i is factor score

4.2.4 Limitation of weighted overlay analysis

The weight selection is still subjective and depends on the expertise of the person involved as well as the available literature. The weighting can also vary considerably depending on the interests who are assigning the weights, which can significantly vary the results. Therefore, the results must be validated with available ancillary data. For example, if groundwater potential zoning map is prepared, then it can be validated with the ground truth data for the region. The validation at certain sites can easily make this analysis available for decision support.

4.3 Overlay operations

The overlay operations involve clipping, erasing or splitting operations. These overlay operations differ from the Boolean methods such that only the geometry and attributes of the data layer are transferred to the output layer unlike Boolean methods where logical operations are considered and attribute data has information from both the layers being considered, as shown in section 4.1. Overlay operations are further explained in this section.

4.3.1 Clip

Clip defines the areas for which features of the data layer will be clipped based on the clipping polygon like an everyday example of cookie cutting. In this case, the attributes of clipping layer are not transferred. The clip overlay operation can be clearly illustrated through Fig 8. The Fig 8a shows a raster data as the data layer and Fig 8b., a vector layer is our area of interest. So, in GIS the area of interest can be easily clipped from Fig 8a, shown in Fig 8c. It can be seen the legend still represents the data of original data layer, Fig 8a. Thus, the attribute information as well geometry for only the data layer is retained. This implies the output layer is still a raster layer, even though clipping layer was a vector data.

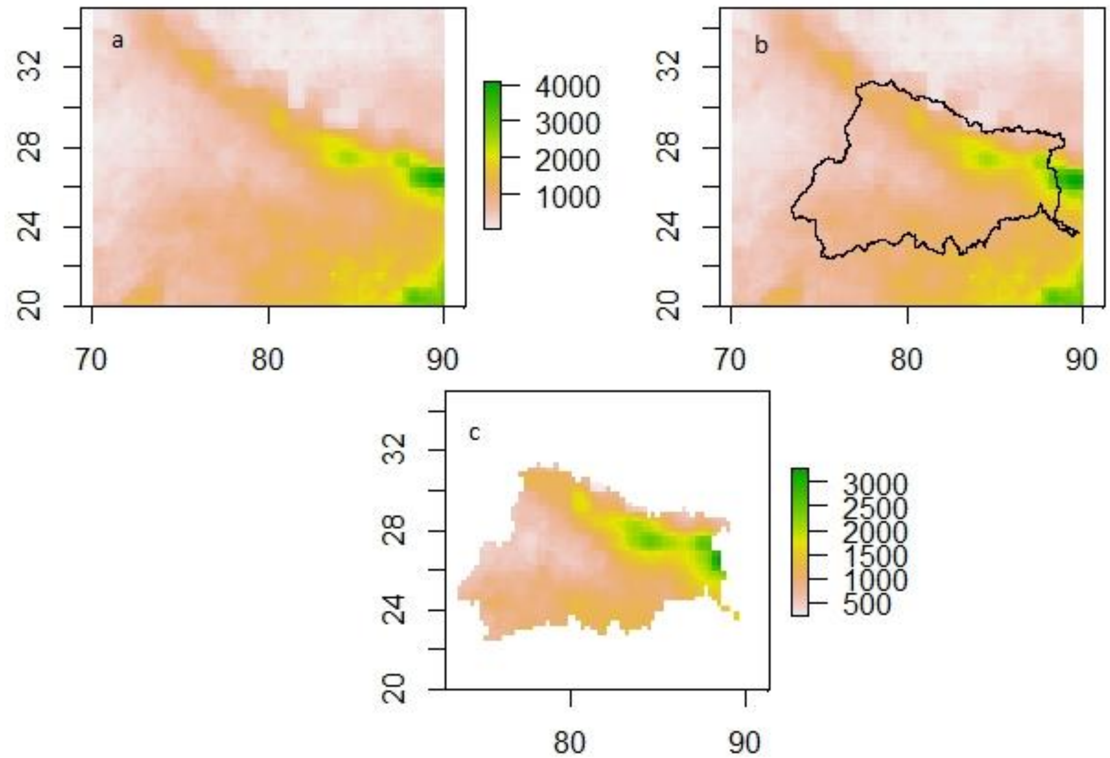

Fig 8. Illustration of Clip overlay operation, a. Data layer, b. Clipping polygon, c. Output layer.

4.3.2 Erase

Erase defines the areas for which features of the data layer will be erased based on the clipping polygon, that is erase is actually opposite of Clip overlay operation. Here also, the attributes of erasing layer are not transferred to the output layer. So, in GIS we can easily erase the area from the layer, which we may not require for further analysis. The erase overlay operation is illustrated through same data layer and polygon in Fig 9, to show the difference between clip and erase overlay operation. The Fig 9a shows a raster data as the data layer and Fig 9b., a vector layer, Fig 9b is our area, which we would want to erase. Fig 9c, is the output layer with attributes of raster layer.

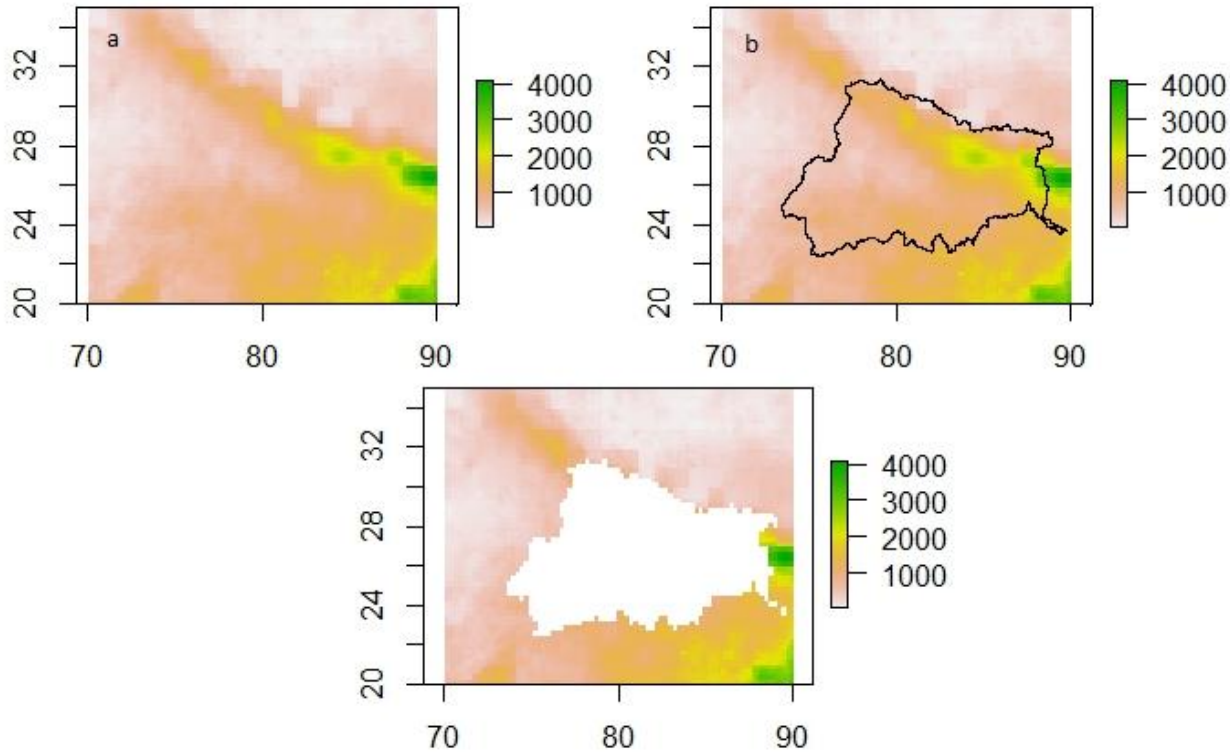

Fig 9. Illustration of Erase overlay operation, a. Data layer, b. Erasing polygon, c. Output layer.

4.3.3 Split

The split operation is used to divide an input layer into two or more layers based on a split layer, shown in Fig 10. The split layer must be a polygon, while the input layers can be both raster or vector. For example, we might want to split soil map based on road map as shown in a simplified splitting layer in Fig 10. Based on the lines, Input layer is split retaining the attributes of split portion of input layer.

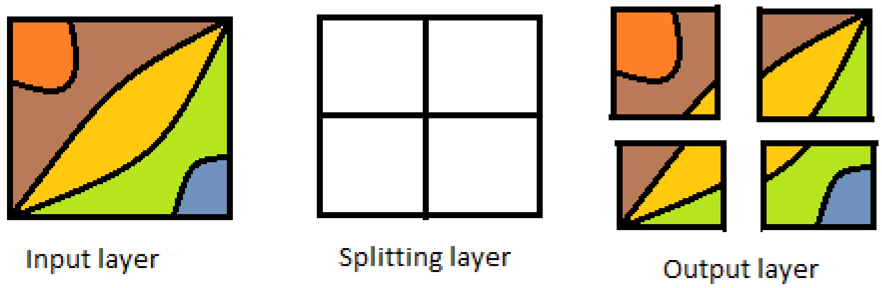

Fig 10. Illustration of Split overlay operation

5. Overlay errors

Overlay method can result in errors, especially when polygon-in polygon overlaying method is utilized for the vector layers. Slivers are a common error produced when two slightly misaligned vector layers are overlain. This misalignment can come from several sources including digitization errors, interpretation errors, or source map errors (Chang 2008). However, this problem is removed using GIS software that incorporate a cluster tolerance option that forces nearby lines to be snapped together if they fall within a user-specified distance.

Another source of error is error propagation. Error propagation arises when inaccuracies are present in the original input and overlay layers and are propagated through to the output layer. These errors can be related to positional inaccuracies of the points, lines, or polygons. Alternatively, they can arise from attribute errors in the original data table(s). These errors can also arise if the overlay operations are being utilized for raster layers only and the resolution of raster layer cell differs. For example, the spatial resolution of TRMM rainfall data is 25 km, while MODIS Land surface temperature(LST) can be obtained at 1km. So, if we overlay the LST layer over TRMM and perform the Boolean operations, there might be error due to positional inaccuracies. Thus, before overlay operations can be performed, the two layers ought to be resampled to the same spatial resolution.

FAQs

Q 1. Explain point-in polygon spatial overlay analysis

Ans. The point-in-polygon overlay determines the points from one input layer lying inside a particular polygon. In this method the output layer is still a point layer but has added attributes of the polygon feature from the second layer. The Fig. 1 shows an example of point-in-polygon, where in point layer shows soil type attribute represented by 1,2 and 3 and second layer is polygon layer, which shows three types of vegetation, namely A, B and C. So, if we want to also know the know the vegetation type associated with the soil type feature overlay is utilized. Thus, here the output layer is again point layer but now the attribute information has additional data representing the vegetation for each point determined by overlaying the two layers. Therefore, the attribute information is changed from 1,2 and 3 to 1A, 2B and 3C respectively as given in Fig.1 in Section, 3.1

Q2. How is point-in-polygon different from line-in-polygon overlay analysis?

Ans: The point-in-polygon overlay determines the points from one input layer lying inside a particular polygon. In this method the output layer is still a point layer but has added attributes of the polygon feature from the second layer. The Fig. 1 in Section 3.1, shows an example of point-in-polygon. Here the output layer is again point layer but now the attribute information has additional data from layer two.

This method is similar to the previous one, only difference being that the input and output layer is now line layer. Here the line feature can be divided into multiple segment depending the polygon it lies within. As shown in Fig. 2, Section 3.2, the output layer is again a line layer but each segment may be divided into additional segment based attribute information of polygon layer. Thus, in point-in-polygon the number of features remain the same and only attributes are added to the point data. However, in line-in-polygon additional feature data is obtained due to division of line segment in various smaller segments based on the polygon of the overlay layer.

Q3. Explain the errors due to spatial overlay analysis

Ans. Overlay method can result in errors, especially when polygon-in polygon overlaying method is utilized for the vector layers. Slivers are a common error produced when two slightly misaligned vector layers are overlain. Another source of error is error propagation. Error propagation arises when inaccuracies are present in the original input and overlay layers and are propagated through to the output layer These errors can be related to positional inaccuracies of the points, lines, or polygons.

Q4. What is weighting by rating approach?

Ans: The principle of not considering each input layer equally and assigning weights according to relative influence level of each layer is known as weighted overlay. Weighted linear sum is one of the approaches that is applied to a raster dataset, grid cells are multiplied by the weights assigned to particular layer and then the summation is done. The weighting can be done according to weighting by rating, where the layers are first ranked and then each layer is given a score.

Q 5. Explain Boolean methods of spatial overlay analysis

Ans: The overlaying methods involve operations like Boolean algebra comprising the logical comparison between the layers. The Boolean algebra method allows the decision making for the categorical data. This overlaying method allows creation of conditional statements, which results in a true/false output layer. The various Boolean methods are Union, Intersect, Exclusive disjunction and Negation. All the four methods are given in text in section, 4.1.

Q6. What is MCE?

Ans: Multi criteria evaluation is weighting method and is based on the Analytic Hierarchy Process (AHP), given by Saaty (1980). The method utilizes a pairwise comparison procedure for the derivation of factor weights. Using a 9-point rating scale, each layer is compared to each other for its relative importance in developing the final solution. The principal eigenvector of this matrix is then calculated to derive the best-fit set of weights. To explain the method, the steps followed in the method are given above in the text in section, 4.2.3.

MCQs

1. In weighting the layers if rank of the layer is y than the weight is $(x\text{-}y+1)^n$, where n determines the weights. What does n=0 imply

   a) weighting is as rank sum

   b) all criteria receive the same weight

   c) neither is true

Ans: b

2. If λ_{max} is 3.0 and number of factors considered is 3, what does CR value imply? (where, RI=0.58)

   a) acceptable

   b) not acceptable

   Ans: a

3. Point-in-polygon spatial overlay analysis can be applied to only

   a) vector data

   b) raster data

   c) both

Ans: c

4. Point-in-polygon spatial overlay analysis adds to output layer

a) feature data

b) attribute data

c) both

Ans: b

5. Weights by rank can be employed using method

I. if rank is y than the weight is x-y+1

II. if rank is y than the weight is 1/y

a) only I

b) only II

c) neither I nor II

d) both I and II

Ans: d

6. Slivers is an error which results due to

a) point-in-polygon spatial overlay

b) line-in-polygon spatial overlay

c) polygon-in-polygon spatial overlay

Ans: c

7. The Boolean operator used for an intersection query is

a) OR

b) NOT

c) AND

d) XOR

Ans: c

8. In which type of overlay does the new layer show only the common features of the input layers?

a) intersect

b) clip

c) negation

d) union

Ans: a

Summary:

Overlay refers to the digital integration of location and attribute information of several spatial layers. Thus, overlay analysis helps in analyzing various input layers and provide new information like a wildfire vulnerability or hazard map. Overlay operations available for use with data include the point-in-polygon, line-in-polygon and polygon-in-polygon models. The methods to combine the various datasets based on Boolean overlay include Union, intersection, difference, exclusion and negation. The more complex method may include the weighting overlay analysis, where in each dataset is not given equal importance. Allocation of weights can vary considerably depending on the interests of the experts who are assigning the weights. and influence the results. The MCE method tries to bring some rational to weighting scheme based on the consistency ratio index calculated based on the method given by Saaty, 1980.

Reference Books:

1. Geographic Information Systems An Introduction, 1990, Prentice Hall, New Jersey
   by Star, J. & Estes, J.
2. Introduction to Geographical Information Systems
   by Kang-Tsung Chung
3. An Introduction to Geographical Information Systems, 2006
   by Heywood, I.; Cornelius, S.; Carver, S.