Subject	Geology
Paper No and Title	**Remote Sensing and GIS**
Module No and Title	GIS Modelling
Module Tag	RS & GIS XXX

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, Faculty of Natural Sciences, Jamia Millia Islamia Delhi	**Dr. Iqbal Imam** Aligarh Muslim University, Aligarh	**Dr. Atiqur Rahman** Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia Delhi

Table of Content

1. Introduction

Modelling is a simplified version of a concept. It is a simplified representation of a phenomenon or a complex system into simple and understandable concept of real world. It is a graphical, mathematical, physical, or verbal representation of a concept, phenomenon, relationship, structure, system, or an aspect of the real world. Therefore, it can be said that modelling is a representation of reality in either *physical* form or *symbolic* form.

A modeling may have following objectives:

- **a)** To facilitate understanding by eliminating unnecessary components,
- **b)** To aid in decision making by simulating 'what if' scenarios,
- **c)** To explain, control, and predict events on the basis of past observations.

Since most phenomenon are very complicated and much too complex, a model contains only those features that are of primary importance to the model maker's purpose.

2. General types of Models

2.1. Structural Model

Structural model focuses on the composition and construction of things. There are two types of structural models:

- **Object Model**: This type of model forms a visual representation of an item. Characteristics include scaled, 2 or 3dimensional, symbolic representation.
  For example: an architect's blueprint of a building.

- **Action Model**: It tracks the space/time relationships of items. Characteristics include change detection, transition statistics and animation.
  For example: a model train along its track.

2.2. Relational Model

Relational model focuses on the interdependence and relationships among factors. There are two types of Relational models:

- **Functional Model**: This model is based on Input / Output method. It tracks relationships among variables, such as storm runoff prediction. Characteristics include cause/effect linkages and sensitivity analysis.
- **Conceptual Model**: It is perception-based. It incorporates both fact interpretation and value weights, such as suitability for outdoor recreation. Characteristics include heuristics (expert rules) and scenarios.

3. GIS Models

When Geographical Information System (GIS) is used in the process of building models with spatial data, it is called as GIS modelling. GIS modelling involves symbolic representation of *Locational* properties (Where?), as well as *Thematic* (What?) and *Temporal* (When?) attributes describing characteristics and conditions with reference to space and time. There are two types of GIS model:

3.1. Cartographic Model: It is automation of manual techniques, which traditionally use drafting aids and transparent overlays, such as a map identifying locations of productive soils and gentle slopes using binary logic expressed as a geo-query.

3.2. Spatial Model: Spatial model is expression of mathematical relationships among mapped variables, such as a map of crop yield throughout a field based on relative amounts of phosphorous, potassium, nitrogen and ph levels using multi-value logic expressed as variables, parameters and relationships.

3.3. Elements of GIS Modelling: A GIS model must have following elements:

- A set of selected spatial variables
- Functional / mathematical relationship between variables.

A model is related to exploratory data analysis, data visualization and data base management. GIS model can be vector based and raster based.

4. Types of GIS Model

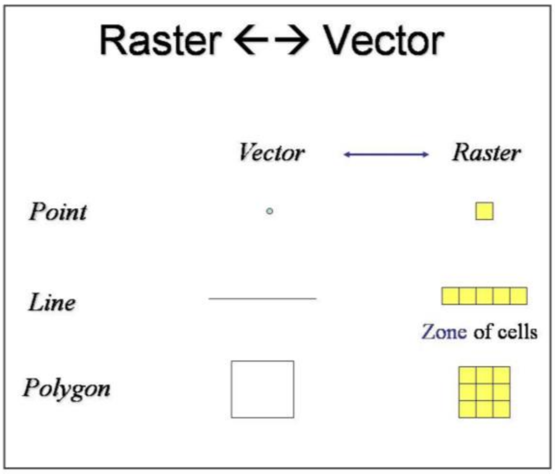
GIS model can be of following types:

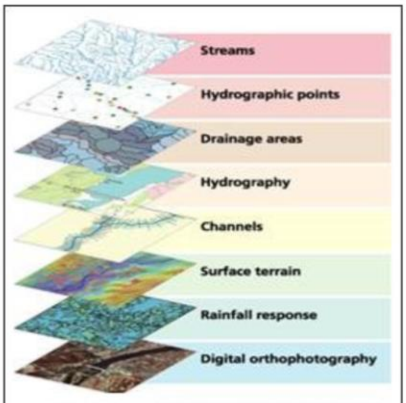
4.1. Binary Models: The Binary Model is a probabilistic information retrieval technique that makes some simple assumptions to make the estimation of document/query. A binary model gives a simple yes (1) or no (0) assessment of a location.

Terms are independently distributed in the set of relevant documents and they are also independently distributed in the set of irrelevant documents. The representation is an ordered set of Boolean variables. More specifically, a document is represented by a vector

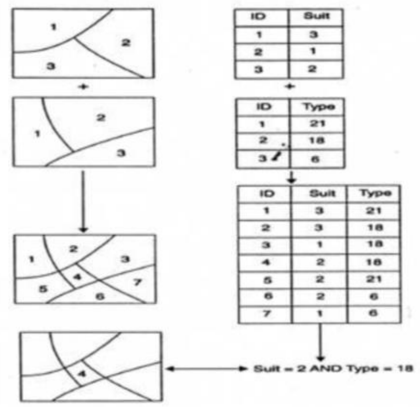
$$d = (x_1, \ldots, x_m)$$

Where, x_t=1 if term t is present in the document d and x_t=0 if it's not. Binary model select features from composite maps or from multiple grids that are presented in the features of point, line, polygon and cell (for raster as well as for vector).

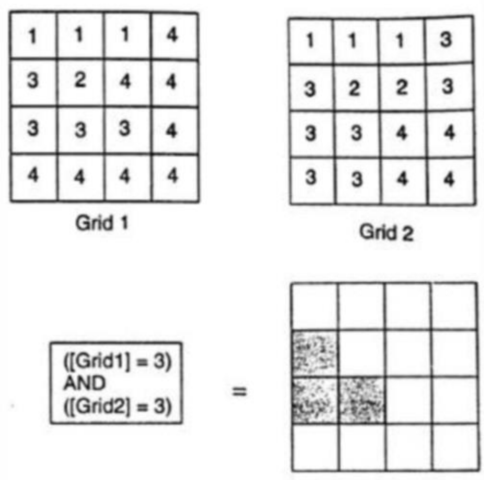




Data model based on inventory of data analysis



An illustration of a vector based Binary (logical) model. The two maps at the top are overlaid so that their spatial features and their attributes of Suit and Type are combined. A logical expression, Suite = 2 AND Type = 18, results in the selection of polygon 4 in the output.



An illustration of a raster based binary (logical) model. A query statement, ([Grid 1] =3 AND ([Grid 2] = 3), results in the selection of 3 cells in the output.

4.2. Index Models (Weight-rating Score Models): In Index model, standardized values are assigned to spatial elements of each layer. Then a rank map is produced by using index value calculated from a composite map or multiple grids.

The index value of a polygon/cell =

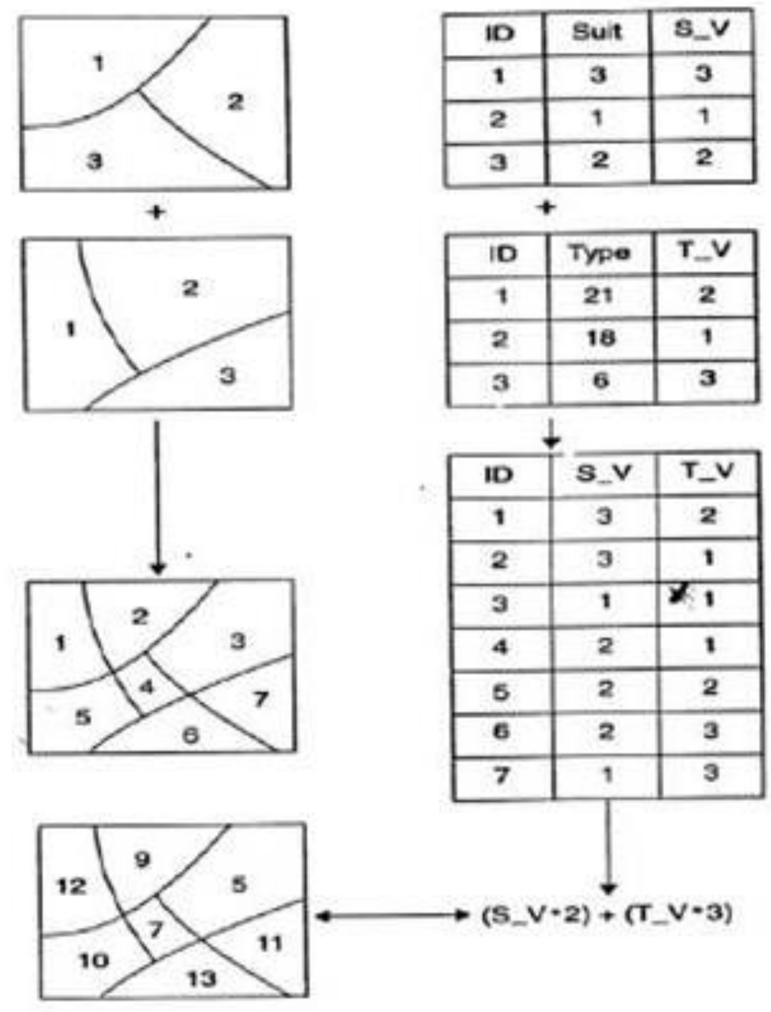
$$\sum_{i=1}^{n} w_i \int i$$

Where, w_i = weight of factor *i* and $\int i$ = rating factor of *i*

It can also be written in this form = (weight$_1$ x score $_{A1}$) + weight$_2$ x score $_{A2}$) + ...

For Index Models following steps are needed:

a) Assign weight to each variable (w).
b) Assign and standardize scores to each class of each variable (data layer)
c) Index value calculation
d) Ranking index values of each polygon/cell.



ID	Suit	S_V
1	3	3
2	1	1
3	2	2

+

ID	Type	T_V
1	21	2
2	18	1
3	6	3

ID	S_V	T_V
1	3	2
2	3	1
3	1	1
4	2	1
5	2	2
6	2	3
7	1	3

An Illustration of a vector based index model.

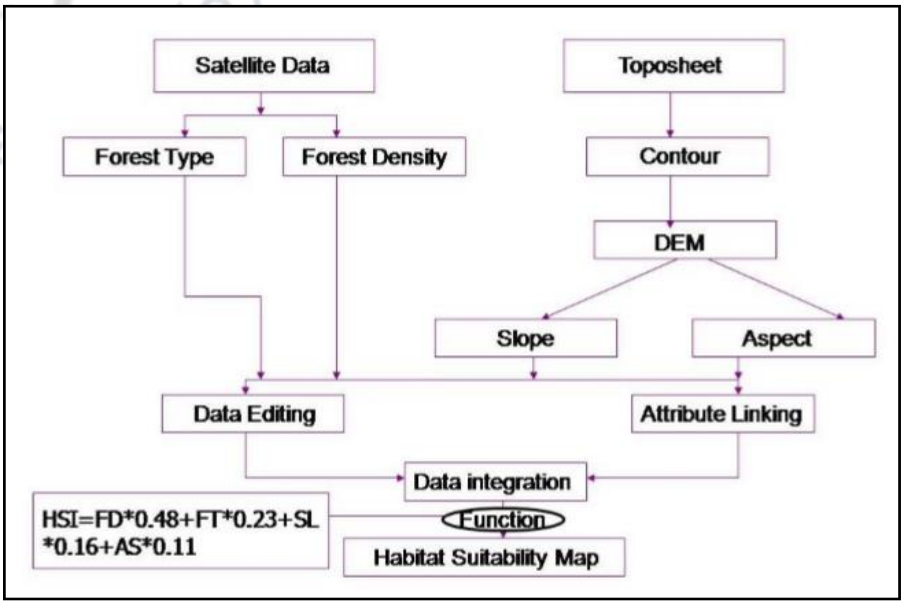
First the observed values of each map are given the numeric scores from 1 to 3. The Suit values of 1,2 and 3 are given the scores of 1,2, 1nd 3 respectively. The type values of 6, 18 and 21 are given the scores of 3, 1 and 2 respectively. Second the two maps are overlaid. Third, a weight of 2 is assigned to the map with Suit and weight of 3 to the map with Type. Finally, the index values are calculated for each polygon in the output. For example, Polygon 4 has an index value of 7 (2*2+1*3)

4.2.1. Examples of Index Modelling: Use of Analytical Hierarchy Process (AHP) along with Remote Sensing GIS for modelling Habitat Suitability Index (HIS)

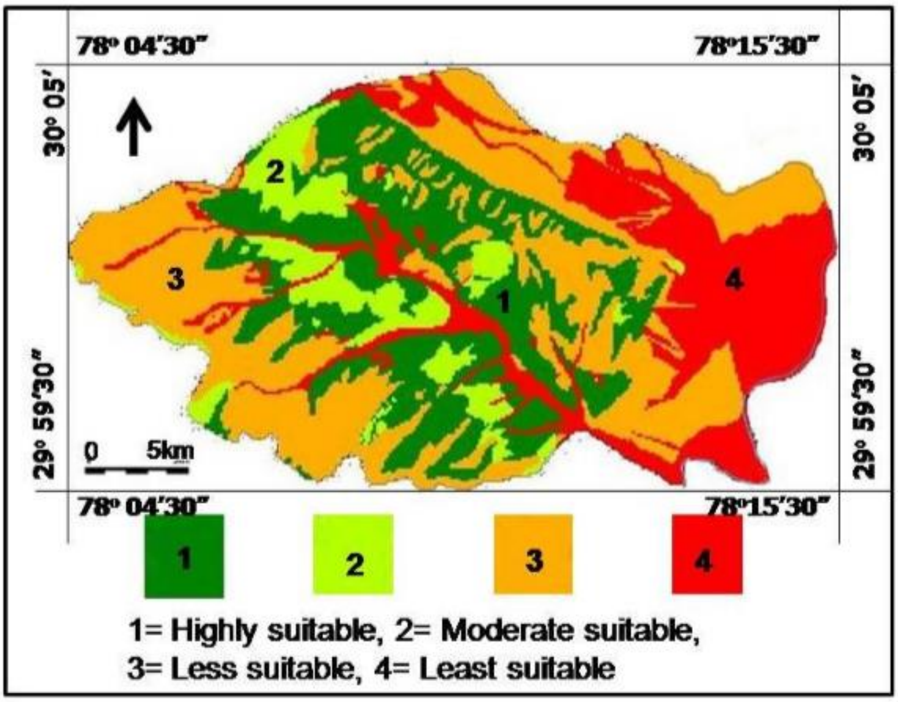
Database Creation: Identification of factors that influence the spatial distribution of animal species is important for habitat suitability index. Therefore, literatures survey on factors influencing tigers' distribution, data from field surveys and suggestions from conservation experts were considered as input data for modelling. Vegetation types, forest density, slope and aspect were selected as variables for developing habitat suitability model. Factors like Forest Types, Forest Density, Aspects and Slope were obtained after analysis of remote sensing imagery in GIS domain.

Topographic maps of study area (scale 1:50,000) were scanned and transferred to ERDAS IMAGINE 8.7 for rectification. From topographic sheets, on screen digitisations of contours (of 30 m interval) were done for generating the digital elevation model (DEM). Further, DEM was used to generate Slope and Aspect maps. Vector files of Forest Type and Forest Density maps of study area were extracted from IRS-1D-LISSIII of 2002 imageries and after rectification, these were transferred to ArcView 3.2 for onscreen digitisation. In the present study Forest Type was categories into four classes of Sal forest, Sal mixed forest, Mixed forest and Scrub & grassland and accordingly attributes were allotted. Forest Density was categories into four classes of >70%, 40-70%, 10-40% and 0-10% and subsequently attributes were allotted accordingly. After preparation of map layers of Forest Types, Forest Density, Aspect and Slope, weight allotment procedure was carried out with the help of specialist and field visit experience.

Identification of weights of factors: In the process of habitat evaluation, identification of relative weights among factors is a primary step. The analytic hierarchy process (AHP) is a decision-making method, which was first derived by Saaty in 1977. It is a combination of quantitative and qualitative processes dealing with complex technological, economical, and socio-political problems. For the advance of providing methodology frame and reducing uncertainty, AHP is widely used in environmental evaluation and regional sustainable management. In this, numerical values are assigned to judge relative importance of each factor. In the construction of pair-wise comparison matrix, each factor is rated against every other factor by assigning a relative dominant value between 1 and 9 to the intersecting cell. In the process of AHP, the prime task of calculation is the eigenvector corresponding to the largest Eigen value of the matrix. Each element in the eigenvector indicates the relative priority of corresponding factor, i.e. if a factor is preferred to another; its eigenvector component is larger than that of the other. A sum/product method is used to obtain the Eigen value and the subsequent eigenvector. The weights finally derived by AHP are used for developing the HSI model.



HSI was calculated as the sum of habitat suitability factors multiplied by corresponding weights determined by AHP. Each reclassified raster layer corresponding to the factors selected, were combined by Raster Calculator function in ArcView to generate the spatial distribution map of HSI. The HSI values in the grid cells are a series of continuous values. To visualize distribution of different levels of habitat suitability index and to facilitate the process of understanding, these values are classified into different classes of highly suitable, moderately suitable, less suitable and least suitable.



4.3. Regression Models: Regression analysis refers to techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Regression models are of following types:

4.3.1. Linear Regression: A spatial regression model can then be used for decision-making. For example, it can answer where suitable locations for police stations are? Spatial regression models are also used to predict future crime locations and even in other cities.

Some of the terminology in regression models:

- Dependent variable (y): What are we trying to predict. (Location of crimes)
- Independent variable (x): Explanatory variables that explain the dependent variable. (Income, education, etc.)
- Beta-coefficient: Weights reflecting the relationship between the explanatory and dependent variable.
- Residual: The value not explained by the model

Regression Formula:

$$y = \beta_0 + (\beta_1 \text{ x } x_1) + (\beta_2 \text{ x } x_2) + ...+ (\beta_n \text{ x } x_n) + \varepsilon$$

Development of Habitat Suitability Index (H.S.I.) is an example of Regression model. It gives an idea about the capacity of a given habitat to support a selected species.

4.3.2. Logistic Regression: Logistic regression is a form of regression which is used when the dependent is a dichotomy and the independents are continuous variables, categorical variables, or both and computer software uses following formula for analyzing the probability:

$$In \text{ (ODDS)} = \left(\frac{\hat{Y}}{1-\hat{Y}}\right) = a + bx$$

Where, $\hat{Y}$ is the predicted probability of the event which is coded with 1 (presence) rather than with 0 (Absence), 1-$\hat{Y}$ is the predicted probability of the other decision, and x is predictor variable.

The logistic regression is used to develop Habitat Suitability Index (H.S.I.) for wild animal also.

Here an example is taken to show the use of logistic regression in developing H.S.I for Gaur (a wild animal).

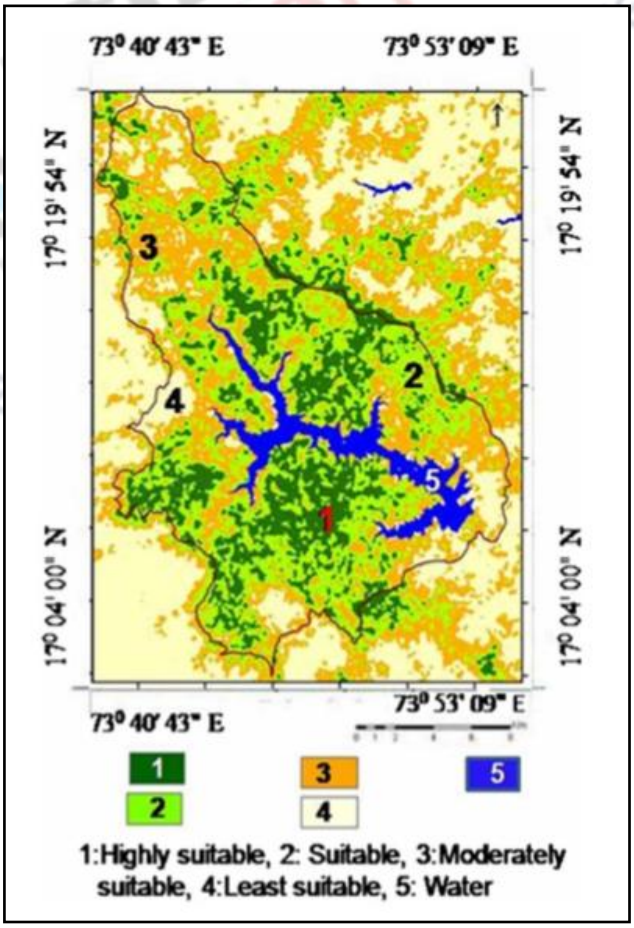
Steps to develop H.S.I. model using logistic regression

- Field visits for collecting data on wild animas' presence/absence. GPS locations of direct or indirect evidences (hoof mark and feeding evidence) of wild animals were marked on FCC. Wherever any direct or indirect evidences of a particular species were found, it was recorded 'presence' for that particular species only and for other species 'absence' was marked. While conducting the field visits ground truthing was also done.
- Create input data base layers of dependent and independent variables. Landuse land cover map, forest crown density map, distance from settlement, road and drainage, elevation, aspect and fragmentation index were taken as variables. These variables were developed after analysing collateral data and satellite imageries of the study area.
- Co-registration of all input layers upto sub-pixel accuracy level.
- After preparation of layer maps, modelling process for H.S.I. for gaur was started. GPS locations of gaurs' presence/absence obtained from the field survey were transferred into ArcView 3.2 and were attached as attributes to all the locations.
- All the independent variables were transferred into raster themes and used for further analysis.
- The points of animal detection were then 'intersected' with all the input layers to produce the habitat use-environmental variables matrix. This worksheet was employed for further statistical analysis. Here, cases of animal sightings were taken as 'Boolean' (presence/absence).
- Logistic regression was run for H.S.I. modelling.
- The coefficients derived from logistic regression were used as weight for variables to integrate all layers in GIS domain to arrive at the probability/suitability map.

$$\mathrm{HSI}=\left[\frac{\{\exp(FT*C)+(FD*C)+(EL*C)+(AS*C)+(RD*C)+(DD*C)+(SD*C)+(FI*C)+(\mathrm{Constant})\}}{\{1+(\exp(FT*C)+(FD*C)+(EL*C)+(AS*C)+(RD*C)+(DD*C)+(SD*C)+(FI*C)+(\mathrm{Constant}))\}}\right]$$

Where, exp = Exponential, DD=distance from drainage, DS=distance from settlement, FD=forest, EL=Elevation, DD=Drainage distance, RD=Road distance, SD=Settlement distance, AS=Aspect, FI=Fragmentation index, C=Coefficient value.

- The estimated log-odds image was then logit transformed to produce the intended probability map. As the log-transform squashed lower values and exaggerates higher values, the classification accuracies had been calculated at cut-off of 0.5. The output map was sliced to ‘not suitable’ at value lower than 0.5 and ‘suitable’ at values higher than that.
- Suitability map was further categorised into four classes of ‘highly suitable’, ‘suitable’, ‘moderately suitable’ and ‘least suitable’.



4.4. Process Models

Next generation of geographic information systems will be driven by process models. These are usually composed of algorithms and heuristics (problem solving technique by trial and error method) that will act on users' requests for the GIS to perform some service for them. For this, it is connected to digital networks to contextualize those requests, and interact seamlessly with other databases and processes to achieve users' goals. Process models are used for following main purposes:

4.4.1. Estimation: The goal of estimation is to determine the value of the regression function for a particular combination of the values of the predictor variables. Regression function values can be estimated for any combination of predictor variable values, including values for which no data have been measured or observed. Function values estimated for points within the observed space of predictor variable values are sometimes called interpolations. Estimation of regression function values for points outside the observed space of predictor variable values, called extrapolations.

4.4.2. Prediction: The goal of prediction is to determine value of a new observation of the response variable. It also determines values (of a specified proportion) of all future observations of the response variable. Predictions can be made for any combination of predictor variable values, including values for which no data have been observed.

4.4.3. Calibration: The goal of calibration is to quantitatively relate measurements made using one measurement system to those of another measurement system. This is done so that measurements can be compared in common units or to tie results from a relative measurement method to absolute unit.

4.4.4. Optimization: Optimization is performed to determine the values of process inputs that should be used to obtain the desired process output. Typical optimization goals might be to maximize the yield of a process, to minimize the processing time required to fabricate a product.

4.4.5. Example of Process Models: Soil & Water Assessment Tool (SWAT): The Soil and Water Assessment Tool (SWAT) is a spatially referenced watershed model used to simulate the impacts of land use, land management, and climate on water quantity. This graphic illustrates the general processes associated with developing and applying SWAT models. The SWAT requires following information as input files and provides results in the form of output files as mentioned below.

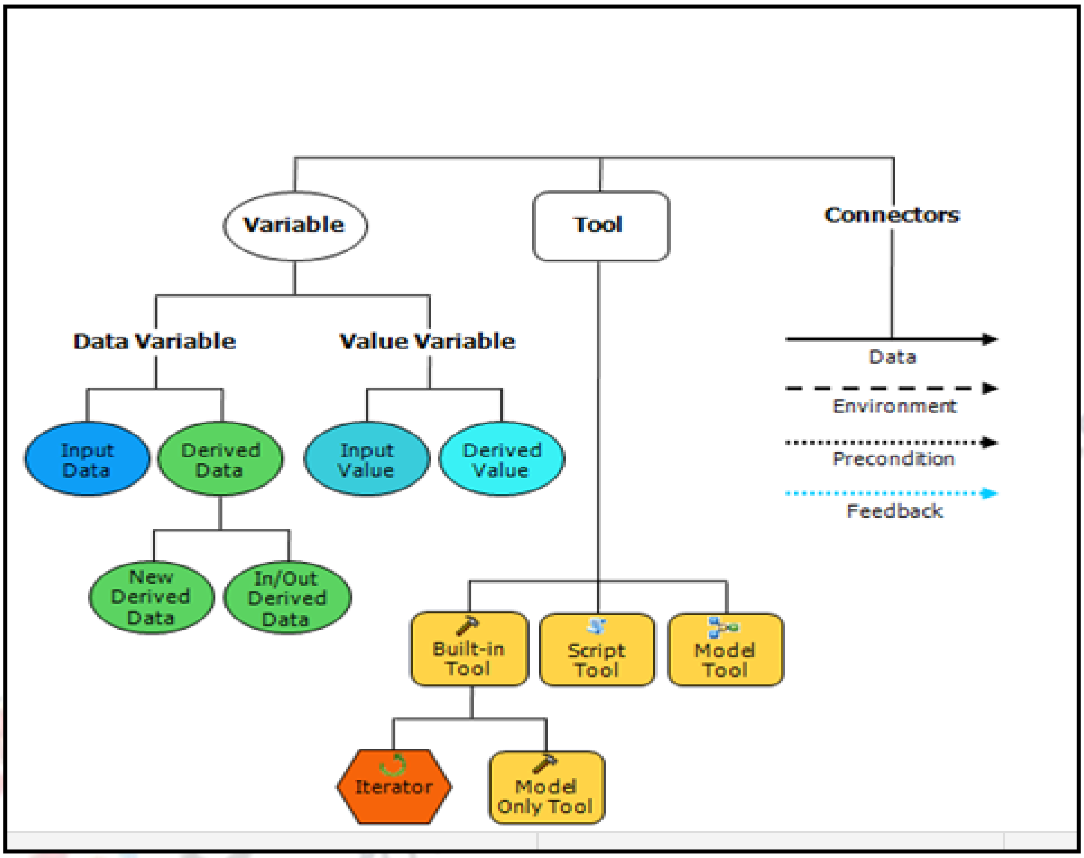
Inputs: Inputs are Land management practice such as Crop rotation, irrigation, fertilizer use, pesticide application rates, and physical characteristics of the basin & sub basin (precipitation, temperature, soil, vegetation & topography).

Output: A number of output files are generated by **SWAT**. These files can be grouped by the type of data stored in the file. There are four output files generated in every SWAT simulation. These files are the standard output file (.std), the Hydrologic Response Units (HRU) output file (.sbs), the sub-basin output file (.bsb), and the main channel or reach output file (.rch). Other files that may be generated include pesticide summary file (.pso), Stream Water Quality Summary File (.wqo), Reservoir Summary File (.rsv), Lake Water Quality Summary File (.lqo). Output data provides simulated values of surface water flow, ground water flow, crop growth, sediment & chemical yields.

WATERSHED MODEL PROCESS
1 MODEL SET-UP – DATA INPUTS
Climate
Land Use
Land Management Actions
Crop Rotations
Cover crops
Filter Strips
Fertilizer Application
Tillage
Irrigation
Wetlands
Drainage
Elevation
Slope
Soil Type
MODEL OUTPUTS
Crop Yields
Amount of Water
Nutrients
Sediments
WATERWAY SCALES
Field
Sub-watershed
River Outlet
COMPARE WITH MEASUREMENTS
2 CALIBRATION
Make informed adjustments of model parameters so model outputs fit observed data, judged by goodness of fit statistics.
3 VALIDATION
Run model with data from another time period or location, testing for statistical agreement between the model output and observations.
WATERSHED MODEL APPLICATION
4 SCENARIO ANALYSIS
Adjust climate, land use, or land management inputs to develop distinct "what if" scenarios to test with the model.
Climate
Past, present, and/or future
Land Use
Agricultural, forested, and/or urban/suburban
Land Management
Implementation extent, intensity, and/or spatial locations of management actions

5. Modelling in ArcMap

A model contains only three things; Elements, Connectors and Text label.

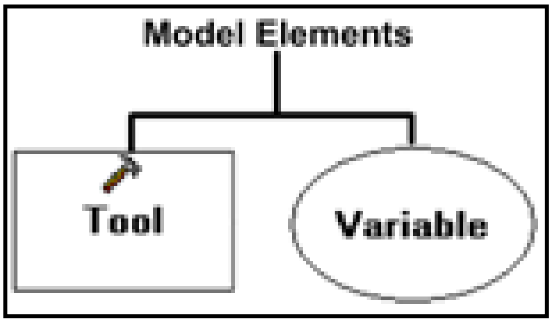


5.1. Elements

Elements are the data and tools with whom we work. There are two types of model elements:

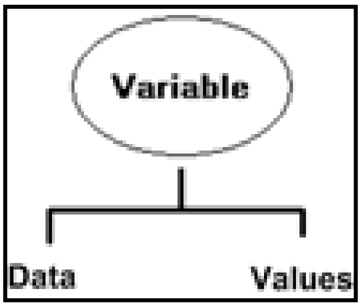
a) Tools and

b) Variables, shown here in an organizational chart:



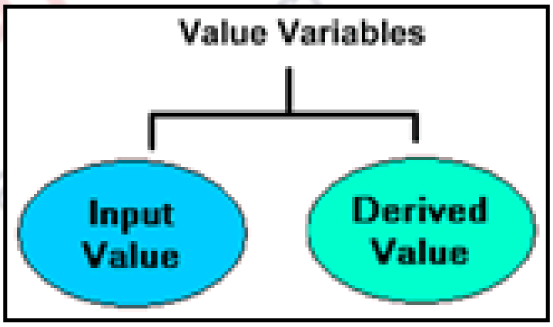
5.1.1. **Tools:** Tool elements are represented with rectangles and are created when we add a tool from Arc Toolbox.

5.1.2. **Variables:** Variables are represented with ovals. It holds values that can be changed. There are two types of variables:



5.1.2.1. **Data Variables:** It reference data on disk or in an in-memory layer (such as a layer in the ArcMap table of contents).

5.1.2.2. **Values:** Values are everything else such as numbers, strings, spatial references, and geographic extents. There are two types of value variables: Input value and Derived value.



5.2. Connectors

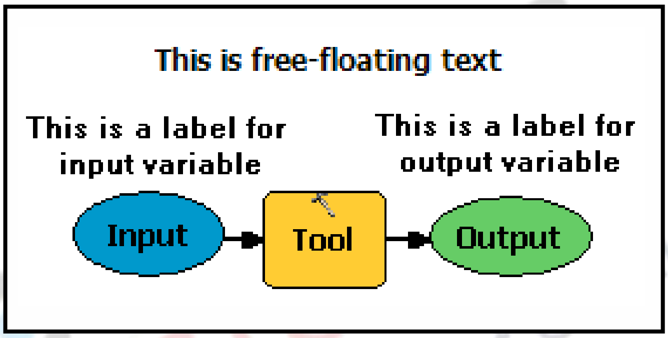
Connectors connect data and values to tools. The connector arrows show the direction of processing. There are four types of connectors:

- **Data:** Data connectors connect data and value variables to tools.
- **Environment:** Environment connectors connect a variable containing an environment setting (data or value) to a tool. When the tool is executed, it will use the environment setting.

- **Precondition:** Precondition connectors connect a variable to a tool. The tool will execute only after the contents of the precondition variable are created.
- **Feedback:** Feedback connectors connect the output of a tool back into the same tool as input.

5.3. Text Labels

In addition to the variable, tool, and connector model elements, there are text label elements, which are graphic elements for explanatory text in a model. A label is not part of the processing sequence. Labels can be attached to elements or float freely in the model diagram.



Frequently Asked Questions-

Q1. **Define modeling in GIS?**

Ans: Modelling is a simplified representation of a phenomenon or a complex system into simple and understandable concept of real world. It is a graphical, mathematical, physical, or verbal representation of a concept, phenomenon, relationship, structure, system, or an aspect of the real world. In other word, we can say that Modelling is a simplified version of a concept. It can also be said that modelling is a representation of reality in either *physical* form or *symbolic* form.

Q2. **What are the different Elements of GIS modelling?**

Ans: A model is related to exploratory data analysis, data visualization and data base management. A GIS model can be vector based and raster based, however, it must have following elements:

a) A set of selected spatial variables

b) Functional / mathematical relationship between variables.

Q3. **Illustrate Process Model with an example of Soil & Water Assessment Tool (SWAT)?**

Ans: The Soil and Water Assessment Tool (SWAT) is a suitable example of process model. SWAT is a spatially referenced watershed model used to simulate the impacts of land use, land management, and climate on water quantity. This graphic, illustrates the general processes associated with developing and applying SWAT models. The SWAT requires following information as input files and provides results in the form of output files as mentioned below:

GEOLOGY	**Paper: Remote Sensing and GIS**
	Module: Modelling in Remote Sensing and GIS

Q4. **What are the objectives a model may have?**

Ans: A model contains only those features that are of primary importance to the model maker's. Therefore, a model must has following objectives:

a) To facilitate understanding by eliminating unnecessary components,
b) To aid in decision making by simulating 'what if' scenarios,
c) To explain, control, and predict events on the basis of past observations.

Q5. **Describe Regression Models?**

Ans: Regression analysis refers to techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Linear Regression, Logistic Regression and Process Models are the examples of regression model.

Multiple Choice Questions-

1. Which of the following components are found in a regression model

a) Independent variables
b) Dependent variables
c) Both a and b
d) None of the above

Ans: c

2. A process model fulfils which of the following purpose (s)

a) Estimation
b) Prediction
c) Optimization
d) All of the above

Ans: d

3. In the process of Analytical Hierarchy Process (AHP) which of the following statement(s) is/are NOT true

a) If a factor is preferred to another, its eigenvector component is larger than that of the other

b) A sum/product method is used to obtain the Eigen value and the subsequent eigenvector

c) In the construction of pair-wise comparison matrix, each factor is rated against every other factor by assigning a relative dominant value between 1 and 9 to the intersecting cell.

d) None of the above

Ans: d

4. When Geographical Information System (GIS) is used in the process of building models with spatial data, it involves symbolic representation of

a) Locational attribute (Where?)
b) Thematic attribute (What?)
c) Temporal attribute (When?)
d) All of the above

Ans: d

5. Which of the following statement is NOT true

a) Modelling is a simplified representation of a phenomenon into simple and understandable concept of real world.

b) It is a graphical, mathematical, physical, or verbal representation of a phenomenon.

c) Modelling is a complex version of a concept.

d) Modelling is a representation of reality in either physical form or symbolic form.

Ans: c

Suggested Readings:

1. Goodchild, M. F. (2003). Geographic Information Science and Systems for Environmental Management. *Annual Review of Environment and Resources*, Vol. 28: pp. 493-519.
2. Granger, K. J. (1990). Process modelling and Geographic Information Systems: Breathing life into spatial analysis. *Mathematics and Computers in Simulation*, Vol. 32(1): pp. 243-247.
3. http://www.colorado.edu/geography/class_homepages/ geog_4203_s08/
4. Hussin Y. A., Matakala M., & Zagdaa, N. (2008). The applications of remote sensing and GIS in modeling forest fire hazard in Mongolia. *In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 37. (Part B8. Beijing): pp. 289–294.
5. Imam E., Kushwaha, S. P. S., & Singh, A. (2009). Evaluation of suitable tiger habitat in Chandoli National Park, India, using spatial modeling of environmental variables. *Ecological Modelling*, 220(24): pp. 3621-3629.
6. Imam, E., & Kushwaha, S. P. S. (2013). Habitat suitability modelling for Gaur (Bos gaurus) using multiple logistic regression, remote sensing and GIS. *Journal of Applied Animal Research*, 41(2), 189-199.
7. Saaty, T. L. (1988). What is the analytic hierarchy process?. *In Mathematical models for decision support* (pp. 109-121). Springer, Berlin, Heidelberg.