Spatial Analysis: an Assessment of the Road Accessibility

Madhu Sudan Adhikari
President
Nepal Surveyors Association (NESA)

Introduction

Spatial analysis is the process of manipulating spatial information to extract new information and meaning from the original geospatial data. Spatial analysis is usually carried out with a Geographic Information System (GIS) because GIS software provides spatial analysis tools for calculating feature statistics and carrying out geo-processing activities like data interpolation and many more. Each user will have different things they are interested in depending on the kind of work they do. Understanding the spatial distribution of data from phenomena that occur in space comprise today a great challenge to the illumination of central questions in many areas of knowledge, be it in health, in environment, in cartography, in geography, in geology, in agronomy, among many others. In hydrology, users will likely emphasize the importance of terrain analysis and hydrological modeling (modeling the movement of water over and in the earth). In wildlife management, users are interested in analytical functions dealing with wildlife point locations and their relationship to the environment. In transportation management, users are interested in the analysis of the accessibility of the people to the road or the transportation network. All of these problems are part of spatial analysis of geo-spatial data. The emphasis of Spatial Analysis is to measure properties and relationships, taking into account the spatial localization of the phenomenon under study in a direct way i.e. to incorporate space into the analysis to be made.

Basic concepts in spatial analysis

Spatial Dependency Principle

Spatial Dependency is a key concept on understanding and analyzing a spatial phenomenon. Waldo Tobler, a pioneer geographer of the University of California, explains this spatial dependency of the geographic phenomenon by his first law of geography, which states that "everything is related to everything else, but near things are more related than distant things". Noel Cressie also explains the nature of the spatial dependency and states, "the spatial dependency is present in every direction and gets weaker the more the dispersion in the data localization increases". It will be contextual to present the few lines of the poem "The Mistress of Version" by Francis Thomson that present the spatial dependency of all the things in the universe:

All the things by a mortal power near or far
Hiddenly to each other linked are
That thou canst not stir a flower
Without the troubling of a star

All the phenomena and occurrences that are happening around us whether they are natural or social, present a relationship among themselves that depends on the distance from each other. This is what we call spatial dependency principle. This principle simply implies that if we find pollution on a spot in a lake it is very probable that places close to this sample spot are also polluted, such pollution decreases with distance, and beyond a certain radius no pollution will be found.

Spatial Autocorrelation Principle

The quantification of the spatial dependency is done by Spatial Autocorrelation Principle i.e. computational expression of the concept of spatial dependence is the spatial autocorrelation. This term comes from the statistical concept of correlation, used to measure the relationship between two random variables. The prefix "auto" indicates that the measurement of the correlation is done with the same random variable, measured in different places in space. We can use different indicators to measure the spatial autocorrelation, all of them based on the same idea: verifying how the spatial dependency varies by comparing the values of a sample and their neighbors'. The autocorrelation indicators are a special case of a crossed products statistics like

$$\Gamma(d) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)\,\xi_{ij}$$

This index expresses the relationship between different random variables as a product of two matrixes. Given a certain distance d, a matrix w_{ij} provides a measure of spatial contiguity between the random variables z_i and z_j, for example, informing if they are separated by a distance shorter than d. Matrix ξ_{ij} provides a measure of the correlation between these random variables that could be the product of these variables, as in the case of Moran's index (Moran's I) for areas, and that can be expressed as

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

Where w_{ij} is 1 if the geographic areas associated to z_i and z_j touch each other, and 0 otherwise.

In these cases the values obtained should be compared with the values that would be produced if no spatial relationship existed between the variables. Significant values of the spatial autocorrelation indexes are evidences of spatial dependency and indicate that the postulate of independence between the samples, basis for most of the statistical inference procedures, is invalid and that the inferential models for these cases should explicitly take the space into account in its formulations.

Statistical Inference for Spatial Data

The consequence of spatial dependence is that statistical inferences on this type of data won't be as efficient as in the case of independent samples of the same size. That is, the spatial dependence leads to a loss of clarifying power. This reflects on higher variances for the estimates, lower levels of significance in hypothesis tests and a worse adjustment for the estimated models, compared to data of the same dimension that reveal independence.

In most cases the more adequate perspective is to consider that spatial data not as a set of independent samples, rather as one realization of a stochastic process. Contrary to the usual independent samples vision, where each observation carries independent information, in the case of a stochastic process all the observations are used in a combined way to describe the spatial pattern of the studied phenomenon.

Data types in spatial analysis

The most commonly used nomenclature to describe the problems of spatial analysis comprises three types of data which are basically found either in vector formats or in raster formats:

Events or point patterns

The point patterns describe the geographic phenomena happening in the reality by articulating the occurrences as points in space. As for example: gasoline stations, police stations, crime spots, disease occurrences, and the localization of vegetal species.

Linear features

The linear features describe the flow or transportation routes of natural or artificial phenomena like road network or transportation networks, electricity transmission lines, river networks etc.

Continuous surfaces

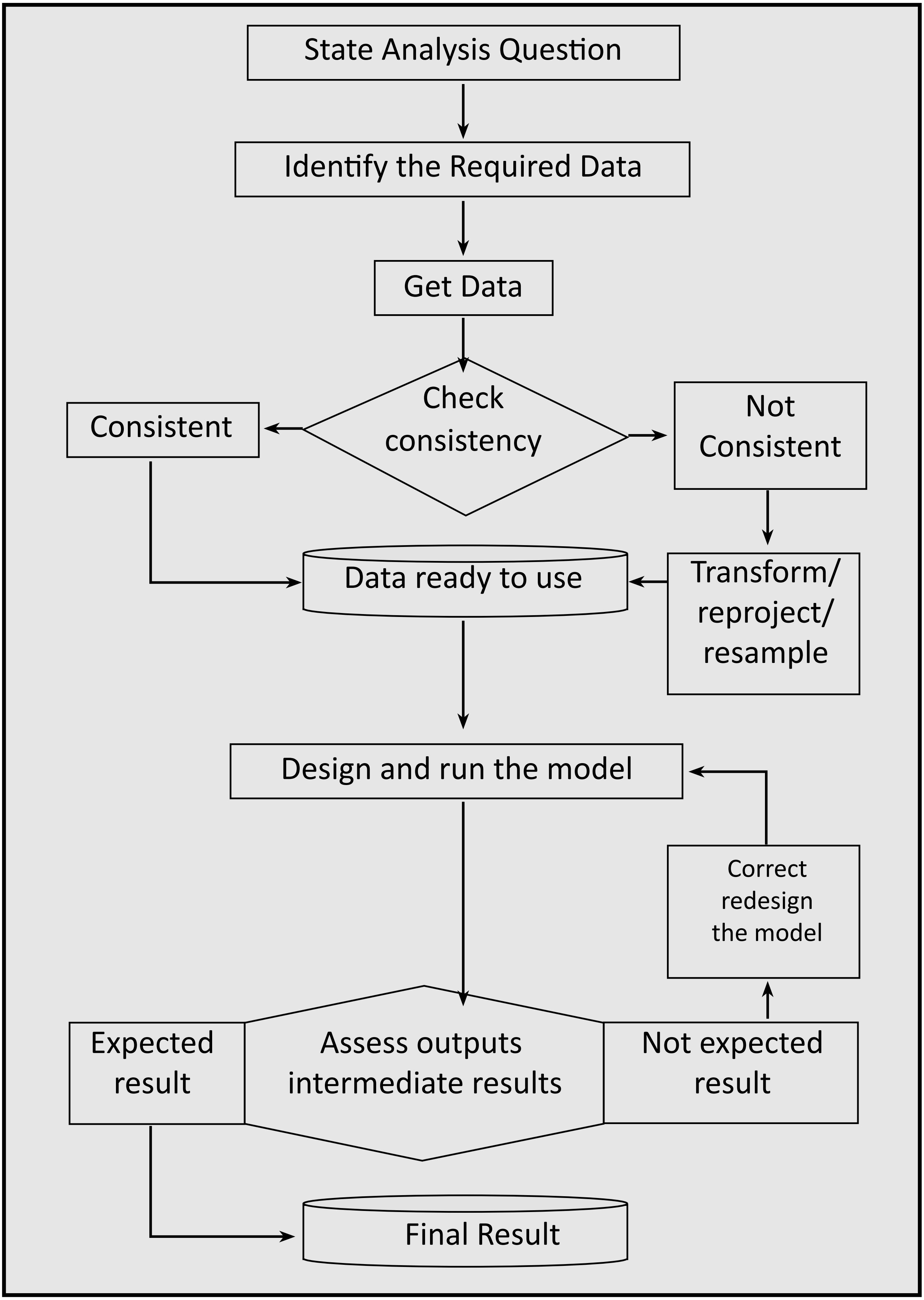
Continuous surfaces are estimated from a set of regularly or irregularly distributed field samples. This type of data is usually obtained from survey of natural continuous phenomena like elevation, surface temperature, barometric pressure, Ph of the surface soil etc.

Areas with Counts and Aggregated Rates

These are the spatial units which are obtained by delineating the areas with the homogeneous data associated to population surveys, like census and health statistics, and that are originally referred to individuals situated in specific points in space. Later these data are aggregated in some spatial units delimited by closed polygons like census tracts, postal addressing zones, municipalities.

Analysis Approach

A systematic analysis approach is very important in the spatial analysis. To understand the analysis question is the foremost and the most important step, it gives ideas to the analyst to model out procedures that lead to achieve the targeted result by processing the available data. Only by understanding the research question the analyst comes to know what data and their formats are required for the analysis and can judge whether the available data and their formats suffice the requirement or not. If the available data do not suffice the requirement the data acquisition and/or conversion is the next step. Another important thing that should simultaneously be taken in to account is the consistency of different data in terms of spatial extent, spatial resolution, and spatial referencing, if they are not they should be made consistent. After getting the data ready for the analysis the model should be designed and run. The intermediate steps should be assessed and the model corrected/redesigned wherever and whenever found necessary. The systematic approach of analysis can be illustrated as following workflow diagram


State Analysis Question
Identify the Required Data
Get Data
Check consistency
Consistent
Not Consistent
Data ready to use
Transform/ reproject/ resample
Design and run the model
Correct redesign the model
Expected result
Assess outputs intermediate results
Not expected result
Final Result

Statement of Analysis Question

Suppose we have the analysis question of the assessment of the accessibility of the road from different parts of the country under the following general assumptions

People can walk four kilometers per hour in flat terrain having slope less than 12%. In the terrain having slope between 12-24%, it is a bit difficult to walk and in such a terrain a person can walk two kilometers per hour i.e. it takes double time to walk the same distance. For the terrain having slope between 24-36% it is very difficult to walk and it takes three times more time to walk the same distance as compared to the terrain having less than 12% slope. For the terrain having slope between 36-48% it is very hard to walk and it takes four times more time to walk the same distance. For the terrain having slope more than 48% it is almost impossible to walk. Only the slope of the terrain is not the hindrance of the travel the river system also obstructs the access to the road. There are some trail bridges to facilitate the traveler and assumption is that the rivers are crossable only from the bridge locations.

Then the main analysis question is that space should be divided in to the regions from where people can access the road with in some fixed period of time. The two general assumptions that determine the cost of the travel time are: the steeper the slope the more costly the travel time and rivers are crossable only from the bridge locations

Identification of the Required Data and Analysis of the Data Consistency

For this analysis we need following spatial data

Slope data (Available in raster of 90 m resolution in imagine image .img format, Modified Universal Transverse Mercator Projection with 84 degree east as the central meridian)

Road network data (Available in poly-line vector in shape file .shp format, Modified Universal Transverse Mercator Projection with 84 degree east as the central meridian)

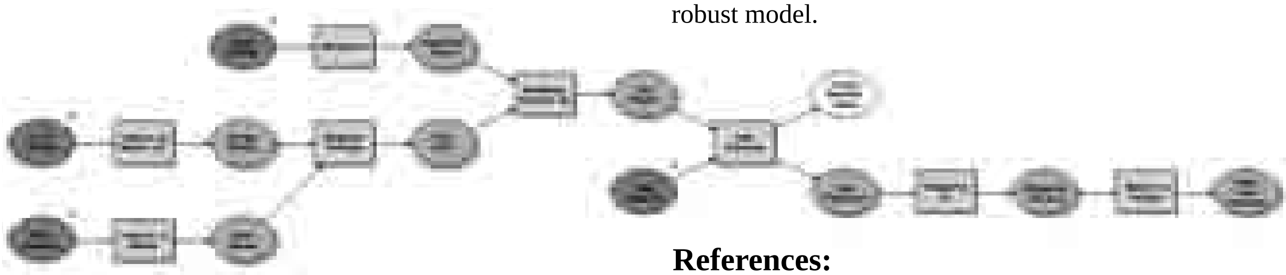
Trail bridge data (Available in poly-line vector in shape file .shp format, Modified Universal Transverse Mercator Projection with 84 degree east as the central meridian)

River network data (Available in poly-line vector in shape file .shp format, Modified Universal Transverse Mercator Projection with 84 degree east as the central meridian)

It is clear that the available data are not consistent for the analysis because the river and the trail bridge the that contribute in the travel costing are in vector format and the slope raster which takes the major stake of the travel cost is in raster format.

Design and Run the Model

For raster based analysis the vector data of the river net work and trail bridge should be converted in to the raster format in the resolution consistent with the resolution of the available slope data. The slope raster should be classified according to its travel cost and it should be overlaid with the other cost contributing raster data of bridge and river in proportion to the corresponding weights and a combined weight raster should be produced from this weighted overlay. The cost distance raster should be calculated from the road network and the just calculated cost raster and the calculated cost distance raster should be classified according to the required time period to access the road by walking.



Result and Discussion:

Dark green belts are the regions from where people can access the road within one hour duration of time where as the yellowish green belt represent the area of the two hour access time. Similarly the yellow red and cyan color belts show the regions of three hour four hour and four-plus hours walk time regions. These regions are determined by the cost distances from the road so they are not parallel to the road and the borders of the regions are not smooth. In the map, the regions 1, 4, 6 and 7 are easily accessible because of the trail bridges where as the regions 2, 3 and 5 are not accessible though they are near to the road because of being in the opposite bank of the river.

Conclusion:

The process of manipulating spatial information to draw new inferences from the original geospatial data is spatial analysis and it is an important component of GIS. Through understanding of the analysis question leads to identify the data requirement and model out the process of geo-computation for the analysis. To make the data consistent in terms of formats, spatial extents, spatial referencing, spatial resolution and overall quality of the expected result is also equally important. The careful assessment of the intermediate and final results guides to design a robust model. In this accessibility assessment experimentation it is clearly seen that all the presumption of the analysis has been taken into account yielding the result as per expectation and the model has been appeared to be a robust model.

References:

1. *Luc Anselin, (1992): Spatial data analysis with GIS; An introduction to application in the social Sciences, National Center for Geographic Information and Analysis, University of California Santa Barbara, CA 93106,*
2. *Chang, Kang-Tsung (2006): Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. (ISBN 0070658986)*
3. *DeMers, Michael N. (2005): Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. (ISBN 9814126195)*
4. *Gilberto Câmara et al, : Spatial Analysis and GIS: A Primer, Image Processing Division, National Institute for Space Research (INPE), Brazil*
5. *Burrough, P.A.; McDonell, R. (1998); Principles of Geographical Information Systems. Oxford, Oxford University Press.*
6. *http://links.esri.com/arcgisdesktop/10/en/resourcecenter/*