

# Data Preparation for GIS based Land Suitability Modelling: A Stepped Approach

Shabir Hussain Khahro<sup>1,\*</sup>, Abd Nasir Matori<sup>2</sup>, Imtiaz Ahmed Chandio<sup>3</sup>, Mir Aftab Hussain Talpur<sup>3</sup>

<sup>1</sup>Department of Engineering Management, College of Engineering, Prince Sultan University, Riyadh, Saudi Arabia

<sup>2</sup>Department of Civil Engineering, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, Perak-Darul Ridzuan, Malaysia

<sup>3</sup>Mehran University of Engineering & Technology, Jamshoro, 76062, Sindh, Pakistan

**Abstract:** The land suitability analysis of any facility has become a complex affair. Each aspect of the landscape has inherent properties that are, to some extent, appropriate or inappropriate for planned activities. Compliance with government guidelines, installation guidelines and customer needs has made this task more complex. Land-use suitability analysis is identifying the most suitable spatial pattern for future land uses according to specific requirements, preferences, or predictors of some activity. Land-use suitability can mean different things to different experts depending on the purpose. Geographic Information System (GIS) has proven to be a key tool in addressing these land compatibility issues. GIS is a science of analysing position information that can count several numbers of these parameters at the same time, and it can be used to identify and analyse earth compatibility for all types of physical planning. A spatial aptitude model generally corresponds to the scientific question "Where is the best location for a particular facility or business? Therefore, this paper aims to provide an intensified approach to prepare datasets and layers for such land suitability problems using GIS. This paper will be valuable to land managers in land-use planning for any facility.

## 1 Background

GIS has provided a new dimension to solve the land suitability (spatial) problems as it has been applied in numerous studies and its success is evident by various researchers [1][2][3][4]. Every feature of the landscape has essential characteristics that are to some degree suitable or unsuitable for the activities being planned. Geographical Information System (GIS) proved to be a key tool to solve such land suitability issues. The results of GIS modelling are often exhibited on a map that is used to focus areas of high to low suitability. A spatial suitability model typically answers the scientific question, "Where is the best location?"[5]. The land suitability analysis becomes a complex activity because it involves various factors, which are difficult to handle, manage and analyse [6][7][8]. Land-use suitability can mean different things to different professionals reliant on the persistence [9][10]. There is a vibrant variance between site selection and site search. Site selection analysis will best identify a particular site for an appropriate activity based on its known potentials, such as location, size, and other attributes. However, a site search problem occurs when there is no determined site for suitability analysis. Due to involvement of numerous factors and mishandling of selection criteria may lead to

improper location selection and GIS made it quite easy as it treats each data factor as a separate layer and in computerized based simulation there are fewer chances of error and bugs as compared with any other manual selection method[8][11]. Hence, this paper aims to highlight the data set collection methods and data layer preparations for such GIS based land suitability problems and modelling. One data set is considered as a case study for this paper. This paper will guide the practitioners working in land suitability modelling to prepare the data sets or layers as per the selection criteria listed for particular site selection for any facility. This approach would help the decision makers to choose the viable location for the development projects and business locations such as petrol filling stations, shopping centres, stadiums, hospitals, industries and others. This approach is a step towards the sustainable decisions of location analysis for future development projects.

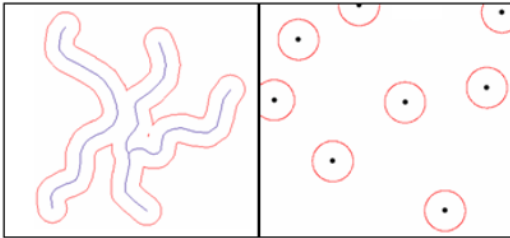
## 2 Development of Layers for Land Suitability Model

The layers are processed through numerous phases before using in model. Following is the stepped approach for data set preparation.

### 2.1 Buffer

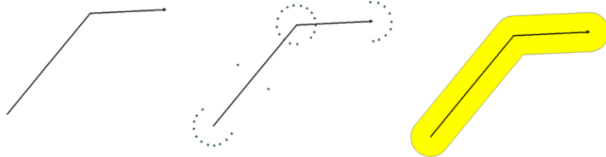
\* Corresponding author: shkhahro@psu.edu.sa

Buffer is the part of proximity tool. It is provided to outline and protect the zones around any specified feature. It provides the clear picture of the areas of influence. It constructs the area feature by spreading outward from point, line or polygon features at certain specified distance. Figure 1 shows the examples of buffered lines and points.



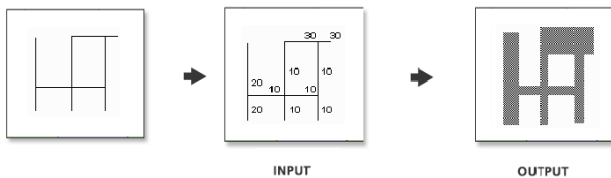
**Fig.1.** Buffering concept of the Point and Line features [12]

The buffer routine traverses each of the input feature vertices and creates buffer "offsets", then the offset creates the buffer features accordingly. Figure 2 shows the buffer philosophy of the following input line feature.



**Fig.2.** Line (1D) feature with Buffer Philosophy [12]

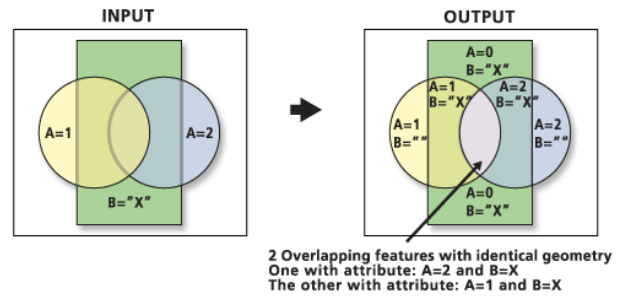
In the above example single valued buffer is placed, whereas for the multi valued buffers, the offsets will be generated according to the buffer distance given to the particular feature. Figure 3 shows the multi valued buffers.



**Fig.3.** Buffering with Multi Variables [12]

### 2.2 Union of the boundary layer

The union tool is used to implement an overlay analysis on a particular feature. This tool creates a new class of functions by combining the features and attributes of each function class. The union normally calculates the geometric intersection of any number of function classes and function layers. Input layers must have a common geometry type, and therefore result in the output function of the same geometry type. It shows that a number of classes of polygonal functions and functional layers can be united together. The union can run with a single input feature layer. In this case, instead of detecting the overlap between the polygon properties of the different functional layers, it will detect the overlap between the single-input functions. Here, the overlapping entities have been separated into new entities with all the attributes of the input function. The zone always overlaps two identical overlapping functions, one for each of the entities involved in the overlap as shown in Figure 4.



**Fig.4.** Union concept of the Data Layers in ArcGIS tool [12]

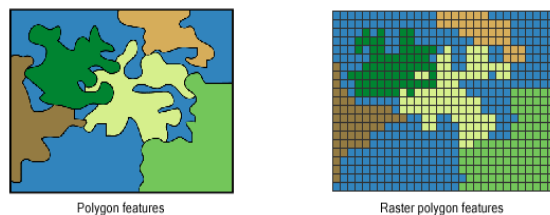
### 2.3 Raster the data layers

After union, in most cases, the data is in vector format then it's better to convert it in to raster format. Raster data format is efficient enough to analyse such land suitability problems. A raster contains a matrix of cells systematized in grid format, where every cell includes a value representing information for particular feature. Data stored in a raster format represents real world phenomena. The level of detail is represented by a raster depends on the cell size. The cell has to be small enough to capture the required detail and large enough to store and analyse the data. Table 1 depicts the detailed characteristics of raster cell.

**Table 1.** Raster Cell Characteristics

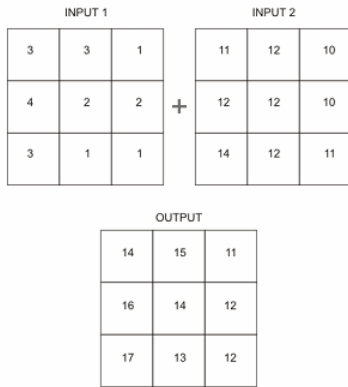
Smaller Cell Size	Larger Cell Size
Better resolution	Poor resolution
Better accuracy	Poor accuracy
Slow display	Fast display
Slow processing	Fast processing
Large file size	Small file size

Figure 5 shows the graphical representation of polygon feature before and after raster command. It can be easy to analyse that this meshed approach will facilitate the overall selection criteria for each data layer.



**Fig.5.** Raster concept in the ArcGIS Tool [12]

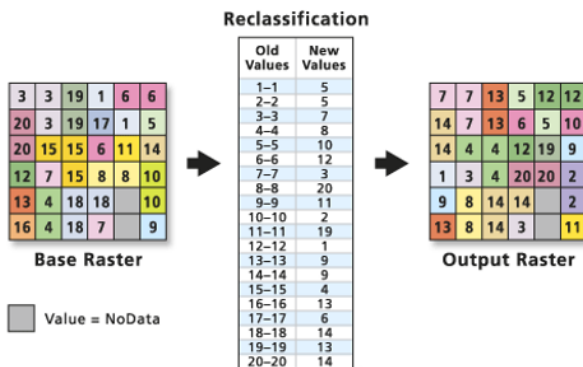
In the raster overlay, every cell in each layer characterizes the same geographic location, which creates possibility to combine the characteristics of different layers in the same layer. In general, numerical values are assigned to each characteristic, which mathematically combines the layers and assigns a new value to each cell in the output layer. Figure 6 is an example of a raster overlay while adding layers. Two input raster layers are added together to create an output raster with the values of each added cell.



**Fig.6.** Raster based Overlay Technique [12]

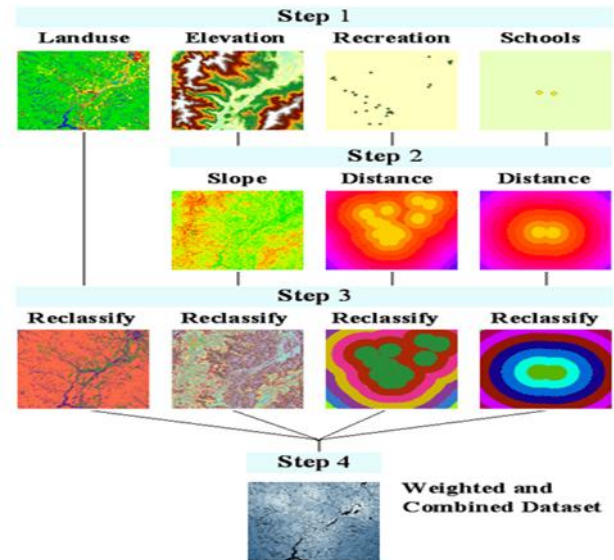
### 2.4 Reclassification

The purpose of reclassification is to change the values of cells to alternative values by different methods. This can be done with one value at a time or groups of values at a time using alternative values based on specific intervals. The reclassification methods are applied to each cell in a zone when an alternative value is applied to an existing value. All reclassification methods use the alternative value for each cell in the source area. No reclassification method uses alternative values for any part of an input field. The Figure 7 shows an example for reclassify the original values from base raster to new reclassified values.



**Fig.7.** Reclassification of the Data Layer in ArcGIS [12]

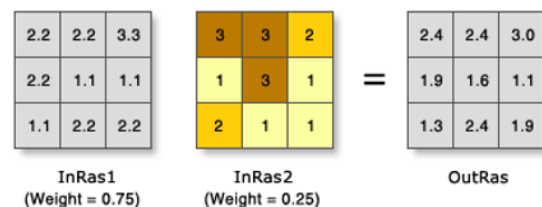
Reclassification is performed to assign preference sensitivity and priority values. The following is an example of an organization chart to find the best location for the school. The input layers are land use, elevation, recreational sites and existing schools. New derived datasets are slope, distance from recreation areas and distance from existing schools. Each raster is reclassified on a scale of 1 to 10. Finally, new reclassified rasters are added to the distance of recreational sites and other schools with greater weight as shown in Figure 8.



**Fig.8.** Spatial Analysis for Weighted Overlay Analysis [12]

### 2.5 Weighted Sum

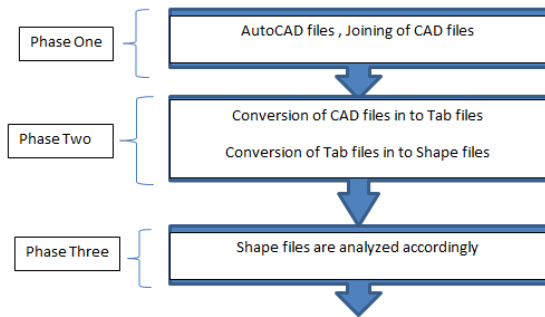
Finally, the weighted sum tool has the ability to weigh and combine multiple input layers to produce an integrated analysis. This corresponds to the weighted overlay tool. In general, the values of the continuous grid are grouped at intervals, e.g. For Euclidean slope or range outputs, each zone is assigned a single value to represent a low, medium, or high importance class. The reclassification tool makes it possible to reclassify such a grid. The weighted sum tool is useful when the floating point output or the decimal weight is required as shown in Figure 9.



**Fig. 9.** Weighted Sum Overlay Approach Concept [12]

## 3 A Case Study for Data Preparation of Single Data Layer

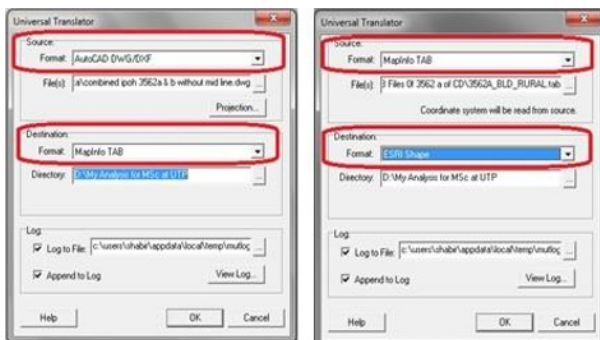
As mentioned earlier, in most cases, the data collected is in the format of dwg, which is supportive to Computer Aided Drawing (CAD) software package. The data is converted into the shape files, which is the acceptable format of ArcGIS software package. The data is processed through different phases to convert it in to shape file acceptable in ArcGIS as shown in the Figure 10.



**Fig.10.** Data Preparation Phases

There are different software packages used to process the data. For phase-I, AutoCAD software package is used to join the data in case same feature data is given in multiple data sets. AutoCAD software package is best supportive to dwg format data for join separate data sets and make single final data set or layer [13].

For phase-II, MapInfo Professional 8.0 software package is used to convert the CAD files into Tab files, meaning that conversion of dwg format files into mapinfo tab file and later on mapinfo tab files are converted into ESRI shape file. The conversion process is shown in the Figure 11.



**Fig. 11.** Conversion Procedure of the Data in “MapInfo Professional 8.0”

In the last phase, ArcGIS software package can be used to analyse the final data layers in the format of the vector shape file. Following steps are used to prepare the final data layers to be used in the model as discussed in detail in earlier sections of this paper.

- Vector shape file
- Placement of the buffers
- Union of the boundary layer
- Defining the projections
- Conversion of the layer from vector format to raster (grid) format
- Reclassification of the layer

To provide a better understanding of the above layer preparation stages, one of the layers is discussed in detail. The vector shape file of the airport data layer is shown in the Figure 12.



**Fig.12.** Vector Shape File of Airport Data Layer

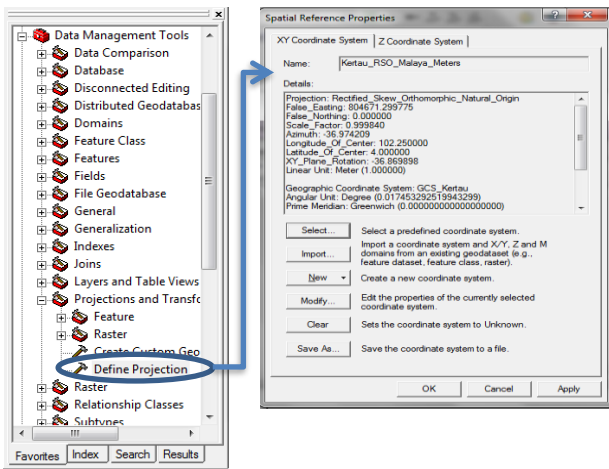
The shaded area represents the airport boundaries in the case study area and the rest is representing the boundary. The boundary may have details of the other features like roads, lakes, rivers, land uses and natural reserves. The next stage is the placement of the buffer distance suggested by the different codes for particular facility. For this example, the airport data layer is provided a buffer of 1500 meters from installation of a new petrol filling station as shown in the Figure 13.



**Fig.13.** Vector Shape File of the Buffered Airport Data Layer

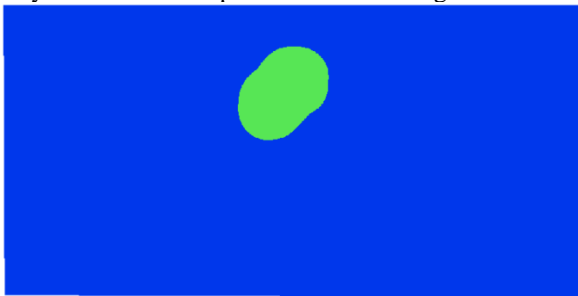
The shaded area represents the airport boundaries with the placement of buffer value. This characterizes that the shaded area is not suitable for the installation of the new petrol filling station. Whereas, the rest area is representing the boundary, which characterizes that this area is suitable for the installation of the new petrol filling station in contrasts to this particular layer. The next stage is the addition of the airport data layer with the boundary layer. Boundary layer is representing the complete case study area. This task is done by using the union command. Before the union, there were two layers, one was the airport data layer and other was the boundary. Now, the data layer will behave as a single data layer and it is easy and robust to convert this vector data layer into raster (Grid-based) layer. This stage is important because it will provide a uniform and a homogenous platform to overlay the different data layers. It is important to display the data layers correctly and thus data layers are provided with a projection. Every data layer has a coordinate system, which is used to integrate it with other data layers. Projections enable the data layers to perform various integrated analytical operations like overlay of the data layers. Although, it is possible to work with the data layers having undefined projection. It is not possible to properly overlay the data layers from different projections without defining a projection first. For this case study, Rectified Skewed

Orthomorphic (RSO) Malaya Grid (meter) is used because it is normally used in Malaysia and West Malaysia. The procedure to define the projection in the ArcGIS software package is shown in the Figure 14.



**Fig.14.** Defining Projections of Data Layer

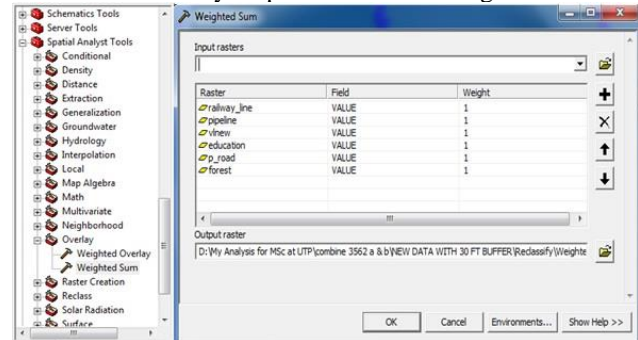
After defining the projections to the data layer, the next stage is to convert the vector data layer into a raster (Grid-based) data layer. This is done by using spatial analyst tool in ArcMap as shown in the Figure 15.



**Fig.15.** Raster Shape File of the Buffered Airport Data Layer

As discussed earlier, raster is a grid based approach. This raster data layer consists of different cells and the cell size is 30m x 30m. The calculation of the suitable and non-suitable land parcels is easy in this case because the whole area is divided into a grid of cells sized 30m x 30m each. The final stage of the data processing is the reclassification of the raster layer. The purpose of reclassify the data layer is to assign numeric values to classes with each map layer, so they have equal importance in determining the most suitable locations [14]. This is done by using the reclassification command of spatial analyst tool. Now the final airport data layer is ready to overlay with the other different data layers. Details of suitable land parcels with reference to the airport data layer can be analysed easily. The blue shaded area in the figure shows the suitable land parcels in perspective of the airport data layer only, which can be used for the installation of new petrol filling stations whereas the green shaded area shows the non-suitable land parcels for the installation of new petrol filling stations. This non suitable zone also behaves like a

constraint. All of the data layers which are required or selected for any particular land use facility are processed accordingly with the placement of suggested buffers. Finally, all the data layers are overlaid to generate the final land suitability map as shown in the Figure 16.



**Fig.16.** Overlay the Data Layers

All the data layers can be overlapped using the overlay command of spatial analyst tool of ArcMap. The calculated weights of all the selection factors can be incorporated here to generate the final land suitability map for any land use facility. It is critical here to share that each land use facility has different set of criteria and this criterion preference and priority can be prepared by selecting any suitable multi-criteria decision making technique. It is mandatory that the total sum of the weights calculated for the selected range of factors should be 1 in decimal valued weights or it should be 100 in ordinary weights.

## 4 Conclusion

Rapid urbanization led to many benefits but at the same time it is quite challenging for decision makers to cope up with enough development activities to meet the demands of customers. Land selection and its suitability became very challenging and imperative today as each land parcel has different characteristic. Hence, the researchers and decision makers have to introduce new methods and approaches to meet these requirements and at the same time meeting sustainability theories which can be enforced by law. Therefore, this paper concludes with one of the possible solution to such land suitability problems. As, the use of geospatial technologies through Geographic Information System (GIS) allow to analyse and interpretation of any land suitability modelling at different scales, time and cost. It also offers any change at any stage easily and allows frequent updates of any land use planning. This stepped approach for data preparation will help the land use practitioners to perform land use modelling easily as it's the predefined validated approach. As, this paper is the part of a full length research for land suitability of petrol filling stations in Malaysia. This paper will also help land use planner in urban land mapping, which is essential to identify and classify areas that are very suitable and less suitable, so that consistent management measures can be proposed and implemented immediately to plan and protect precious lands in a sustainable way. It can also be

extended for monitoring and planning any future development of an area. It is also beneficial for land managers and planners to develop regional plans at any scale.

## Acknowledgement

The authors are thankful to **Prince Sultan University, Riyadh, Saudi Arabia** for support and expertise role throughout this research study. The authors also extend their gratitude to Universiti Teknologi Petronas, Malaysia and Mehran University of Engineering & Technology, Jamshoro, Pakistan.

## Reference

1. J. Malczewski, "GIS-based multicriteria decision analysis a survey of the literature," *Int. J. Geogr. Inf. Sci. Vol. 20, No. 7, August 2006, 703–726*, 2007.
2. S. H. Khahro, A. Matori, I. A. Chandio, M. Aftab, and A. Talpur, "Proposed GIS-based Environmental Impact Assessment approach for site suitability of Petrol Filling Stations in Malaysia," 2012.
3. Z. LASHARI, M. Y. MANGI, N. SAHITO, and ..., "Land Suitability Analysis for Public Parks using the GIS Application," *Sindh Univ. Res. Journal...*, vol. 49, no. 3, pp. 505–512, 2017.
4. I. A. Chandio, A. N. B. Matori, K. B. Wanyusof, M. A. H. Talpur, S. H. Khahro, and M. R. M. Mokhtar, "Computer Application in Routing of Road using Least-Cost Path Analysis in Hillside Development," *Res. J. Environ. Earth Sci.*, vol. 4, no. 10, pp. 907–911, 2012.
5. C. G. Njoku and A. O. Alagbe, "Site suitability assessment of Petrol Filling Stations (PFSs) in Oyo Town, Oyo State, Nigeria: A Geographic Information Systems (GIS) approach," *IOSR J. Environ. Sci. Toxicol. Food Technol.*, vol. 9, no. 12, pp. 8–19, 2015.
6. J. W. Mugo, P. C. Kariuki, and D. K. Musembi, "Identification of Suitable Land for Green Gram Production Using GIS Based Analytical Hierarchy Process in Kitui County, Kenya," *J. Remote Sens. GIS*, vol. 5, no. 3, 2016.
7. M. M. Aburas, S. H. O. Abdullah, M. F. Ramli, and Z. H. Asha'ari, "Land Suitability Analysis of Urban Growth in Seremban Malaysia, Using GIS Based Analytical Hierarchy Process," *Procedia Eng.*, vol. 198, no. Supplement C, pp. 1128–1136, 2017.
8. Shabir Hussain Khahro and Zubair Ahmed Memon, "GIS based land suitability analysis using AHP model for urban," *Int. J. Civ. Eng. Technol.*, vol. 8, no. 10, 2017.
9. D. I. A. Syed Maheena, Khan Junaid N., Kumar Rohitashw, "Land suitability for paddy and maize in a watershed using remote sensing and Geographic Information System(GIS)," *J. Soil Water Conserv.*, vol. 16, no. 4, 2017.
10. Z. G. Rahmat *et al.*, "Landfill site selection using GIS and AHP: a case study: Behbahan, Iran," *KSCE J. Civ. Eng.*, vol. 21, no. 1, pp. 111–118, 2017.
11. A. A. El Baroudy, "Mapping and evaluating land suitability using a GIS-based model," *Catena*, vol. 140, pp. 96–104, 2016.
12. B. Booth and A. Mitchell, "Getting Started with ArcGIS GIS by ESRI."
13. Z. J. Pudlowski, U. International, and E. Education, "Monash University's Gippsland Campus' innovative going green-based system for infrastructures education with environmental cultural change applications," *World Trans. Eng. Technol. Educ.*, vol. 6, no. 2, 2007.
14. C. D. Tomlin, *Cartographic Modelling*. 1990.