Evaluation of Indexing Overlay, Fuzzy Logic and Genetic Algorithm Methods for Industrial Estates Site Selection in GIS Environment

Hamid Ebadi, Roozbeh Shad, Mohamad Javad Valadanzoej,Alireza Vafaeinezhad

Faculty of Geodesy and Geomatics Eng. K.N.Toosi University of Technology
No 1346, Mirdamad cross, Valiasr st., Tehran, IRAN
Tell: (21) 877 0218
Fax: (21) 878 6213

1-Assistant professor, Email: ebadi@kntu.ac.ir
2-Graduate student, Email: Rouzbeh_Shad@yahoo.com
3-Assistant professor, Email: valadanzouj@kntu.ac.ir
4-Graduate student, Email: Arvafaei@Noavar.com

Commission V,WG V/2

KEY WORDS: GIS, Site Selection, Model, Indexing, Fuzzy, Genetic

ABSTRACT:

One of the main parameters which is helpful for industrial development in each country is land use mapping. With respect to land use mapping and industrial development strategies, managers as decision-makers can organize the best location for industrial estates manufacturing. There are enormous data volume and complex criteria for the site selection of industrial plants that cause much more difficulty for decision making. Accordingly, by the use of GIS as information technology and its analytical capability for decision making optimization, we can overcome these difficulties. GIS analytical functions can answer “What is the optimum solution?” with respect to the GIS users’ requirements. Overlay is one of the spatial functions that can combine spatial data layers from diverse resources for the site selection applications using integration models. Integration models, based on their implementation methods, are divided in some groups (for example: Boolean operation, Indexing overlay, Fuzzy logic, Genetic algorithm, Weight of evidence, etc.). In this case study, at first, we selected a study area that was convenient for our purpose which was located at the north-west of Iran. Then, effective parameters and criteria were defined for industrial estate site location and corresponsive data layers. Finally, we classified and prepared data layers with respect to main criteria and parameters. By checking the executive routines for different kinds of integration models, we evaluated results of Indexing overlay, Fuzzy logic and Genetic algorithm methods that could be implemented in GIS environment based on the processing time and spatial accuracy which presented some interested models for industrial estate site location.

1 INTRODUCTION

Nowadays, in each country economic and sustainable development are related to industry and mine capacity and type of used technology. Industry and mine development with optimum concentration are caused more facilities on the social lifeway. Since the beginning of twentieth century, industrial constructions in forms of industrial zone, region and estate have been considered for countries industrial development (Poladdezh, 1997). To achieve this purpose, it is necessary for each country attending to land use mapping problems and work force distributions and locating convenient sites for industrial constructions. This subject depends on some factors such as population, employment, land use, environment, etc.

Using GIS as an information technology and efficient spatial decision making tool, industrial estates factors managing and analysing will be done better. In this paper, we will try to find optimum solution for industrial estates site selection and applying solution in GIS .

2 SITE SELECTION

Site selection is locating convenient sites with introducing efficient criteria and factors using some integration models.

2.1 Criteria and Factors

There are some effective parameters and conditions that influence on the site location for special application. These parameters are extracted of collected spatial data layers in study area and prepared for entering to integrated models using some data processing methods (Bonham Carter and G.F., 1991).

2.2 Data Processing Methods

Data Processing consists of some operational activities that are performed on unprocessed raw data for entering GIS analytical functions. Each processing method has one or some outputs that depends on processing method type, entering raw data structure (Vector or Raster) and requiring data for entering analytical functions. There are different kinds of processing methods in GIS environment (for example: data combination, vector to raster conversion, distance map production, reclassification, etc).

2.3 Integration Models

Various models were used for real world events simulation in GIS environment (Aronof, 1989). Integration model is one of them that used for site selection by integrating related spatial data and effective criteria. There are various integration models that categorized by kind of functions and their executive routine (Knowledge driven or Data driven).

Knowledge driven: Experts experience and scientists science are used for executing models.

Data driven: Models are executed based on existent solutions and dependency value computation.

We'll describe some useful models that have been used in our application and mention our reasons for selecting them. These models consist of: Boolean Operation, Indexing Overlay, Fuzzy Logic and Genetic Algorithm.

2.3.1 Boolean Operation : In this model, input maps are integrated by using logical operators such as AND, OR, XOR and NOT (Bonham Carter and G.F., 1991). Although, Boolean Operation is an easy and fast model to run, there are some problems in its execution routine. In this model, all input factor maps (processed data) have same weights and appropriate sites can’t be separated based on their priority. In our research for industrial estate site selection, this model isn't used because of mentioned problems and weighted nature of efficient criteria and conditions in our application.

2.3.2 Index Overlay Model: Factor maps are integrated using following Equation :

$$S = \frac{\sum W_i S_{ij}}{\sum W_i} \quad (1)$$

where W_i= The weight of ith factor map
S_{ij}= The ith spatial class weight of jth factor map
S= The spatial unit value in output map

Comparing Index Overlay with Boolean Models executive routines, it is identified that Index Overlay model has more flexibility and ability for priority indication on spatial units of factor maps (Bonham Carter and G.F., 1991). With respect to mentioned characteristics, this model is useful for comparing and evaluating integration models in industrial estate site selection process.

2.3.3 Fuzzy Logic Model: In industrial estate site selection application, for each factor map we can define classes and spatial units as a subset that their membership values in industrial estates convenient location are mapped between 0 and 1. There are some fuzzy operators such as Fuzzy AND, Fuzzy OR, Fuzzy Product, Fuzzy Sum and Fuzzy γ that are used for factor maps integration (An and Moon, 1991).

Fuzzy AND : This operator operates like common statements in classic sets theory as Equation 2.

$$\mu_{Combination} = MIN(\mu_A, \mu_B, \mu_C, ...) \quad (2)$$

where
$\mu_{combination}$= Each spatial unit value in output map
$\mu_{A,B,C,...}$= Spatial units membership values

This is used when there have been two or more factors or evidences together that can help solving the problem. In our application, this operation isn't used because of its weakness on participating all effective factors and lack of specific evidences for industrial estates location.

Fuzzy OR : This operator operates like union statements in classic sets theory as Equation 3.

$$\mu_{Combination} = MAX(\mu_A, \mu_B, \mu_C, ...) \quad (3)$$

where $\mu_{combination}$= Each unit value in output map
$\mu_{A,B,C,...}$= Spatial units membership values

This operator is used when there are sufficient positive factors and evidences in study area. In our evaluation, because of lack of positive factors for industrial estates site selection, Fuzzy OR operator isn't used.

Fuzzy Product : Fuzzy Product Operator produces input factor maps membership values and emplaces results on output map (Equation 4). Therefore, it has decreasing affects on results and is used when input factor maps debilitate each other.

$$\mu_{Combination} = \prod_{i=1}^{n} \mu_i \quad (4)$$

where $\mu_{combination}$= Each unit value in output map
μ_i= The weight of ith factor map

In industrial estates site selection, There is not weakener factor maps then fuzzy product isn't used in our evaluation.

Fuzzy Sum : This operator is defined as following Equation.

$$\mu_{Combination} = 1 - \left(\prod_{i=1}^{n} (1 - \mu_i) \right) \quad (5)$$

Where $\mu_{combination}$= Each unit value in output map
μ_i= The weight of ith factor map

This operator is used when input factor maps have increasing effects in each other. In our application, because of increasing effects of Accessing factors and Infrastructure parameters, this operator was used and compared with other integration operators.

Fuzzy γ : This operator is a general form of Fuzzy Product and Sum operators (Equation 6).

$$\mu_{Combinaticn} = (Fuzzy\ Sum)^{\delta} * (Fuzzy\mathrm{Produet})^{1-\delta} \quad (6)$$

According to Equation 6, γ can change from 0 to 1. This operator is used when there have been increasing and decreasing effects on interactions of evidences (Zimmermannand and Zayo, 1980). This operator is used for models evaluation because we don't know about any factor maps interactions such as Environment Parameters and natural factors.

2.3.4 Genetic Model: Before describing Genetic execution routine, it is necessary to introduce some corresponsive concepts as below.

Choromosom : It displays one possible solution for a required problem. In site selection problem each set of selected spatial units are primary solutions for industrial estates location and indicates one Choromosom. Each unite in mentioned set is called Gene (Gen, 1997).

Population : The possible amounts of Choromosoms in each generation, correspond to the population. On the other hand, population is a set of solutions (Gen, 1997). In industrial estates site selection, all selected spatial units of corresponsive layer (primary solutions) are called initial population.

Fitness Degree : With comparing population of new and previous generation, fitness degree is determined using fitness function. This parameter identifies surviving feasibility in the next generation for all of choromosoms (Gen, 1997). For Genetic model execution in industrial estates site selection chromosomes spatial units values in the new generation are compared with corresponding spatial units values in previous generation. Then, fitness function is Algebraic comparing statement that identifies strong genes for industrial estate construction by cutting poor genes out that aren't accordant to function.

Control Parameters : They are specific conditions such as number of generations, number of choromosoms, length of choromosoms and model terminating condition that caused upper efficiency on model execution routine (Gen, 1997).

In Genetic model first, initial population are selected. Then, factor maps are integrated based on definite integration operators (such as index overlay) and second generation is generated. After that, with using fitness degree, poor genes are appointed and eliminated. Finally, new generation is selected as initial population and above procedure is repeated until reaching constant state. We can decrease repetitions by entering control parameters in the execution routine.

Therefore, this model is selected for our evaluation because site selection is a complex and composite problem with different criteria and conditions and genetic algorithm is a good model for solving these kinds of problems.

3 EXECUTION AND EVALUATION OF MODELS

In this section, we describe regular steps that identify the optimum model for industrial estates site selections as shown in following flowchart.

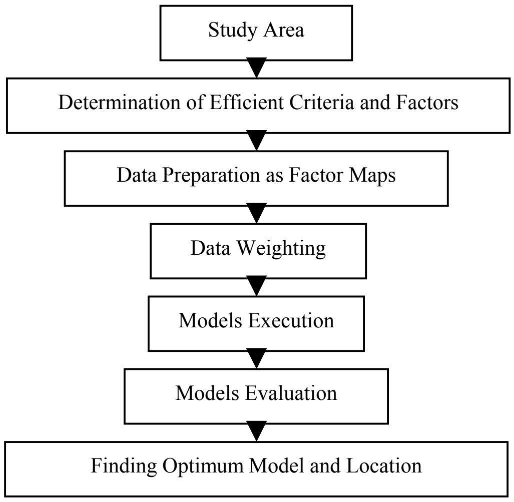

Figure 1. Models execution and evaluation steps

3.1 Study Area

Study area characteristics affect on industrial estates factors and criteria. More features diversity and spatial data density increase factors and criteria. For example, political factors are important on boundaries of each country. In the case of political boundary of study area, we usually consider political conditions. Therefore, its complete recognition causes better selection of effective factors and criteria.

This area is selected based on some parameters such as economic factors, politic strategies, data limitations and land physical attributes (Demers, 1992). According to experts' ideas of Iranian industrial estates company and necessary factors, first, we selected Azarbaijan province at the north-west of Iran. Then, a zone that covers a region of 1959.59 (km^2) area and located around Tabrize (capital of the east Azarbaijan province) was used as our study area. According to Figure 2, there are good density of data in our selected study area.

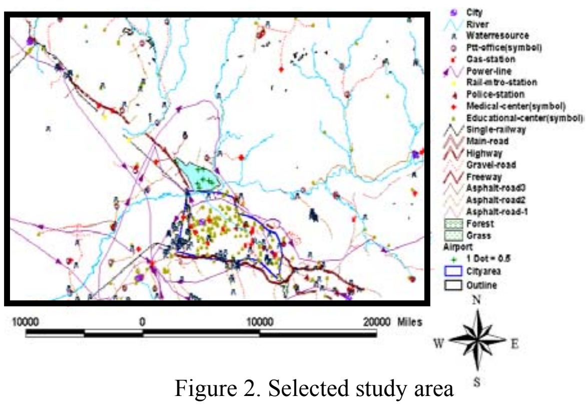

Figure 2. Selected study area

This area consist of different GIS ready layers with scale of 1:25000 produced by NCC (National Cartographic Center of Iran) and edited to be GIS ready data by K.N.Toosi University of Technology.

3.2 Efficient Criteria and Factors

By checking study area properties, available data layers and industrial conditions for industrial estates site selection, criteria and factors are classified in four main groups as below.

Accessing Factors: Layers consist of Freeway, Highway, Asphalt Road 1, Asphalt Road 2, Asphalt Road 3, Gravel Road, Railway, City, Educational Centre, Medical Centre, Police Station, Train Station, Village and Airport.

Infrastructure Parameters: Layers consist of Power Transformation Line, Water Supply, Oil and Gas Station, Telephone, Telegraph and Post Office.

Environment Parameters: Layers consist of Urban Area, Forest and Green Space.

Natural Factors: Layers consist of Slope Map, River and Wind Direction.

3.3 Data Preparation

Before entering data into analytical models, they must be prepared with respect to models execution routine and required inputs used by processing methods. There are many data processing methods in GIS environment, but we selected four of them such as Layers Combination, Data Structure Conversion, Distance Map Supplying and Data Classifying. Major parameter that affects on processing methods selection is data input structure (Bonham Carter and G.F., 1991). Raster structure was selected as input structure in our application because of its properties such as simplicity and calculation ability for integrating layers, and mentioned models (Section 2) execute routine. In here, there is a major point called the optimum data volume capacity that with respect to the scale of maps and required accuracy can be determined. Required accuracy in our application was determined to be 0.3 millimeter on the map scale using experts ideas and required content map. Therefore, maximum pixel size will be 7.5 * 7.5 meter on the ground.

Accordingly, data processing methods were applied on raw data and resulted factor maps. Each pixel were applied on the factor maps has a gray value that indicates amount of proportionality for industrial estate construction with comparing other pixels value or spatial units. For producing factor maps we must do regular steps using processing methods that depend on required criteria and factors. For example there is a vector structure Wind Direction layer of Natural class that must be converted to raster form using Data conversion method and classified based on spatial data priority using Data Classifying method. This priority is determined based on expert ideas and spatial units value. In the wind direction layer, spatial units that their direction are toward city, take less gray value than others because of controlling air pollution in the city.

3.4 Data Weighting

The weight of each factor map indicates amount of its cost and value as comparing with the other factor maps. Correct weights can help finding convenient location for industrial estate area. There are two ways for weighting factor maps that are called Knowledge Driven and Data Driven Weighting.

Knowledge Driven Weighting: In this method, data are weighted in a definite range using expert experience about application. First, different ideas are collected and their dimensions are uniformed. Then, weights are normalized in the defined range using an appropriate scale.

Data Driven Weighting : Weight of each factor map will be determined by calculating amount of dependency value between factor maps and specified sites layer. One of the appropriate methods for determining dependency value is Weight of Evidence. This is a probabilistic method that can be used for integration of factor maps. This model is executed based on probability theory and uses preferent and recent probability rules similar to Baysian model (Bonham Carter and G.F., 1991). There are two parameters in the weight of evidence method that are named sufficiency and necessity ratio. They specify amount of each factor map sufficiency and necessity for industrial estates location using follow Equation.

$$LS=\frac{P(B_i \mid D)}{P(B_i \mid \overline{D})} \quad (7) \qquad LN=\frac{P(B_i \mid \overline{D})}{P(\overline{B}_i \mid \overline{D})} \quad (8)$$

where B_i= The effective zone in each factor
$\overline{B}_i$= The supplemental zone of B_i
D= The existent sites location
LN= Necessity ratio
LS= Sufficiency ratio

Then, by Naperian logarithm computing LN and LS, positive weight of evidence and negative weight of evidence for each factor map can be determined (Bonham Carter and G.F., 1991). Finally, by computing constant parameter that called factor map contrast, dependency value is determined (Equation 9).

$$C=|W_+-W_-|=|Ln(LN)-Ln(LS)| \quad (9)$$

Where W_+= positive weight of evidence
W_-= negative weight of evidence
Ln= Naperian logarithm
C= factor map contrast

Table 1. Weighting of factor maps

Class	Factors	Weight	Class	Factors	Weight
Access Factors 0.27	Main Roads	0.10	Environment Parameters 0.20	Forest	0.35
	Asphalt Roads	0.09		Green Area	0.20
	Gravel Road	0.07		City area	0.45
	Railway	0.12	Natural Factors 0.24	Slope Map	0.45
	City	0.09		River	0.20
	Education Centre	0.07		Wind Direction	0.35
	Medical Centre	0.07	Infrastructure Parameters 0.29	Power Line	0.25
	Airport	0.13		Water Resource	0.25
	Police Station	0.06		Post Telegraph and Telephone Office	0.25
	Train Station	0.10		Oil and Gas Station	0.25
	Village	0.10			

Data driven weighting has less blunders than Knowledge driven, but its correct operation depends on primary existent sites accuracy. In our evaluation, because of inaccuracy locations in the primary industrial estates layer, we used data and knowledge driven weighting simultaneously. The results of data weighting are shown in Table 1.

3.5 Execution of Models

At first, weighted factor maps must be integrated by using integration models. For this purpose, four programs were written and user interface was customized for running Index Overlay, Fuzzy Sum, Fuzzy γ and Genetic models.

Index Overlay execution : This model was executed in two stages. First, in each class, factor maps were integrated with respect to Table 1 and were resulted in four class factor maps. Then, output factor maps were integrated using designed interface.

Fuzzy Sum execution : Fuzzy Sum model was executed based on Fuzzy Sum operator described in section 2.3.2 similar to Index Overlay model.

Fuzzy γ execution : Fuzzy γ model was executed based on its operator described in 2.3.2. Important problem in its operator was determining correct γ. For this purpose, we wrote an auxiliary program that for each input γ it produced output map and computed dependency value between output map and existent industrial estates layer using Equation 9. Then, Convenient γ was determined by changing γ value between range 0 to 1 and comparing dependency values (Figure 3).

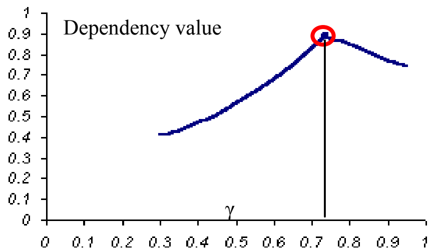

Figure 3. Dependency Value diagram

According to Figure 3, the best γ occurred in 0.74 with maximum dependency value. Finally, factor maps were integrated by entering γ =0.74 and running Fuzzy γ program.

Genetic execution : In this model, first, primary solutions zone was defined. Then, Fitness function was introduced as below.

$$|Gw_{(i)}-Gw_{(i-1)}|<= 0.174 \qquad (10)$$

where $Gw_{(i)}$= The weight of genes in the new generation
$Gw_{(i-1)}$= The weight of corresponding genes In the previous generation

In Equation 10, 0.174 was determined based on subjective experiences and amount of dependency between factor maps and primary selected solutions.

3.6 Evaluation of Models

In this stage, models execution results are evaluated and the optimum model is determined. Performance and accuracy are two important parameters that usually influence on GIS models execution routine. Model performance depends on operation of reference programs algorithm that is used for running model and entering data volume capacity. Because of similarity of entering data volume capacity and using one programmer and programming language, running time was selected for evaluating models execution routine. Then, running time was computed using a precise chronometer and doing accurate observations (Table 2).

Table 2. Running time for each model

Model	Running Time(sec)	Model	Running Time(sec)
Index overlay	36.33	**Fuzzy γ**	196.64
Fuzzy Sum	194.39	**Genetic**	128.61

According to Table 2:

- Index Overlay has the least running time with comparing to other models. The reason for this can be originated from its operator linear operation.
- Fuzzy models are slower than Index Overlay because their operators have more complexity calculation and non-linear properties.
- Genetic model is slower than index overlay because its algorithm depends on entering factor maps in each repetition step.
- Fuzzy γ is slower than Fuzzy Sum because of its less computing operations.

Therefore, Index overlay has the best running time for industrial estates location. It's necessary that mentioned, in MIMGIS software, it is important that analysis have had minimum running time value because of MIMGIS software is run on the network workstations and clients must execute models in the minimum executing time.

Accuracy of models is another parameter that is important for selecting the optimum model. Therefore, we need a criterion for comparing accuracy of output model results. In this case study, existent industrial estates were selected as comparing criterion and their location appropriateness were determined. By paying attention to experts' ideas for each existent industrial estate in study area, status location value was determined based on its annual efficiency and position affects on this parameter. Status location value was defined as very good (A), good (B), bad (C) and very bad (d). After that, with respect to experts ideas each output map was classified in four classes as A (0.7-1), B (0.5-0.7), C (0.3-0.5) and D (0-0.3). Then, by comparing location status value and output maps estimated status, each correct estimating was identified by "+" and each incorrect estimating was identified by "–" (Table 3).

Table 3. Evaluation results for models accuracy

Status Location		Index Overlay		Fuzzy Sum	
Estate	Status	Status	Evaluation	Status	Evaluation
Rajaee	**A**	**A**	+	**A**	+
Salimi	**B**	**B**	+	**A**	-
Saeed Abad	**B**	**B**	+	**A**	-
Status Location		Fuzzy γ		Genetic	
Rajaee	**A**	**B**	-	**B**	-
Salimi	**B**	**B**	+	**C**	-
Saeed Abad	**B**	**C**	-	**C**	-

It is seen in Table 3, that Index Overlay has better accuracy than other models. Therefore, we propose Index Overlay as an optimum model for industrial estates site selection.

3.7 Finding Optimum Location for Industrial Estates in Study Area

With respect to models evaluation result, optimum location for new industrial estate construction can be located by analyzing Index Overlay output map (Figure 4).

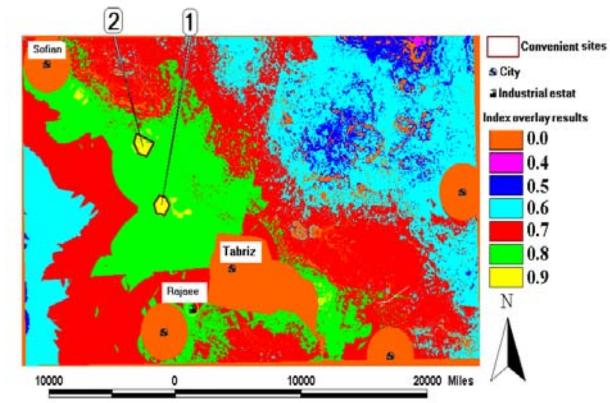

Figure 4. Finding optimum location for new industrial estate

According to Figure 4, convenient sites have been represented by yellow area and the maximum appropriateness (0.9). They are potential areas that decision must be made on them based on their priority (Table 4).

Table 4. Potential sites properties

Sites	Area (km^2)	Distance Form Tabrize (km)	Loction
1	2.72	11	North-west of Tabriz
2	3.59	17	North-west of Tabriz

Based on table 4, it is identified that Site no. 1 has better economic, accessing and facilities conditions than site no. 2, because it is nearer to Tabriz. Price of land is also cheaper in site no. 1 than that of site no. 2. Therefore, we propose site no.1 as an optimum site around Tabriz for new industrial estate construction.

4 CONCLUSIONS

Land use mapping is a convenient tool for optimum relationships adjusting between humans activities and environment effects that is used for logical resources utilizing and managing. Optimum site selection is an important problem for industrial goals that is related to industrial land use mapping. Industrial estates site selection is one of the industrial goals that is very important for each country industrial development. Therefore, in this case study we tried to prove that GIS is able to work as an efficient decision making tool for industrial land use mapping problems by successful running and implementing industrial estates site selection models. For this purpose, first, we selected a study area at the north-west of Iran and evaluated different integration models by entering effective criteria and factors and running models in GIS environment. Finally, by executing optimum model we found optimum location for industrial estate construction.
Weighting of factor maps is very important in site selection process that output results accuracy depends on its correct defining.
By checking different integration models, it was identified that Index Overlay, Fuzzy Sum, Fuzzy γ and Genetic models can be used for industrial estates site selection, although, Index Overlay has optimum running time and output accuracy comparing with other models.

5 REFRENCES

References from Journals:
An, P., Moon, W.M., 1991. "Application of Fuzzy Set Theory for Integration of Geological, Geophysical and Remote Sensing Data.", Canadian Journal of Exploration Geophysics,Chapter 27: PP. 1-11.

Royce, W.W., 1970. "Managing the Development of Large Software Systems.", Proceeding IEEE Wescon, PP. 1-9.

References from Books:
Aronof, S., 1989. Geographic Information Systems: A Management Perspective, Ottawa, Canada: WDL Publications.

Bonham Carter and G.F., 1991. Geographic Information System for Geoscientists: Modelling with GIS, Pergamon, Ontario, PP. 319-470.

Demers, M.D., 1992. Fundamental of Geographic Information System, Chapter 16, PP. 389-401.

Gen, M., 1997. Genetic Algorithms and Engineering Design, John Wiley, Newyork, PP. 5-63.

Poladdezh, M., 1997. Site selection for industrial projects, Tehran, Iran, Chapter 4.

Zimmermann, H.J. and Zayo, P., 1980. Latent Connectives in Human Decision Making: Fuzzy Sets and Systems, Chapter 4: PP. 37-51.