2. LITERATURE REVIEW

2.1 Spatial Information System

The pattern analysis in the agroecosystem analysis procedure is an important process for participants in such an exercise to present their disciplinary and specialist knowledge (Conway, 1986). Spatial patterns are most readily revealed by maps. Overlays are particularly useful in uncovering potentially important functional relationships. Thus in the Chiang Mai Valley, Gypmantasiri, et al. (1980) indicated that cropping intensity was determined by the form of irrigation systems rather than by soil types.

Spatial data analysis is a multidisciplinary concern. Geographic, sociological, agricultural, resource management and earth science activities among others require spatial data analysis (Guptill, 1989). Spatial data sets are frequently heterogeneous, having data elements such as soils, land use, and population statistics, and are often comprised of data sources with different scales, coordinate systems, accuracies, and aerial coverage.

A GIS database is often conceptualized as a series of thematic categories or topics (layers) of information held within the database. These layers may contain information that has been captured from aerial photography, remote sensing satellites, conventional maps, or other sources. The GIS database has two components, locational or geographic, and accompanying feature attribute databases. Locational databases are usually digitized from maps. Attribute databases identify what the features represent in the form of numeric or textual information (e.g., a soil type, road class). GIS packages are specifically designed to store, retrieve, analyze, and display spatial data sets of such characteristics.

There are at least two fundamental ways of representing topological data (Burrough, 1986). Raster representation with set of cells located by coordinates and vector representation with three main geographical entities, points, lines, and areas. In the PC ARC/INFO package, a vector based GIS, its has an ARC module to handle coordinate of the features, while its INFO module handles the descriptions features and how each feature relates to each others (ESRI, 1990).

Database design is part of the system development process. Designers first gather database requirements through user interviews, examination of reports and forms, analysis of existing computer files and databases. A model of data is built during requirements analysis (Hawryszkiewycz, 1990). The model defines major component of user data. These components are the data about the major system entities together with data that describe the interactions or relationships between these major entities. Data modeling is needed to clearly specify the user data structure and to produce a model that is understood by both users and computer professionals.

A data model is a representation of data, relationships, operations, and rules. The highest level of abstraction is the conceptual one, where data is represented by entities, relationships, and attributes from the point of view of the system. Data modeling organizes the data into the optimal form for manipulation, flexibility, and stability. Data modeling is data-driven as opposed to process-driven.

It involves identifying the data entities which the system is concerned with the relationships between those entities, and the attributes of the entities and their relationship.

An entity is an object which information needs to be maintained. The object could be a person, place, or event. An entity must have properties that can be described well enough to distinguish one entity from another. Associations that exist between entities are called relationships.

After defining Entity-Relationship (E-R) model, construction of a set of normalized relations is made. This set of relations removes any unnecessary redundancy. Then the logical database definition is created. The form of the definition depends on the database management software that is to be used to implement the database. Entity-Relationship modeling usually begin by identifying all such major entities and grouping entities with the same properties into entity sets.

Spatial database is a collection of spatially referenced data that acts as a model of reality (Goodchild, 1990). Spatial database functions parallel those of a non-spatial database management system (DBMS), but with extensions beyond the addition, deletion, revision, and Boolean retrieval capabilities of a standard DBMS (Guptill, 1989). The database systems provide the means of storing a wide range of such information and updating it without the need to rewrite programs as new data are entered.

The use of GIS approach can be traced to the need to measure the area of land resources to reclassify and dissolve prior to display, and to overlay data sets and to compare them spatially. GIS technology is of considerable interest in land management. Examples are the work of Davis and Dozier (1990) in Southern California and Susilawati and Weir (1990) for forest land management in Indonesia. Another resource management project was the Columbia River and Tributaries Irrigation Withdrawals Analysis Project by Johnson et al. (1982). Recent interest in automated land evaluation system (Rossiter, 1990; Burrough, 1989; and Wang et al., 1990) also demanded GIS as an integrated part of the system.

2.2 Land Evaluation

2.2.1 FAO Framework

Land evaluation is a process of assessment of land performance when used for specified purposes, involving the execution and interpretation of surveys and studies of landforms, soils, vegetation, climate and other aspects of land in order to identify and make a comparison of promising kinds of land use in terms of applicability to the objectives of the evaluation (FAO, 1976). The FAO's method contains several key innovations. First is the land utilization type (LUT), the LUTs are the land use options, and include in their definition all aspects of land use that can be affected by differences in land. The land qualities (LQ) are attributes of land that directly influence its suitability for one or more uses. They are more abstract than land characteristics (LC), which are the attributes of land that can be measured or estimated. LQ may be edaphic (e.g. nutrient availability), climatic (e.g. radiation regime), agronomic (e.g. soil workability), geographic (e.g. location), or related to hazards (e.g. erosion hazard). The land use requirements (LUR) are the condition of land necessary for the successful and sustained practice of a given LUT. A final innovation of the FAO's method is the concept of land mapping units (LMU), which are identifiable areas of the earth's surface, considering all properties of relevance to land use options (Rossiter, 1990).

The evaluation based on the FAO's method requires voluminous data and are tedious if many possibilities are to be evaluated (Fig. 1). Manual procedure, both for construction of matching tables or transfer function and for calculation of suitability, are time-consuming and error prone (FAO, 1984 and Rossiter, 1990). Different methods have been used to assign suitability classes to land mapping units. Some of them are;

2.2.1.1 Law of Minimum

Evaluation is made by matching the measured or estimated values or classes of land characteristics against the range of requirements listed for each land utilization type. The most limiting rating out of the land characteristics in a single quality is taken as the rating for that quality (FAO, 1983).

2.2.1.2 Multiplication Method

FAO (1983) also used multiplication method for rating suitability classes. In this method, a land quality (LQ) was rated according to its suitability with the value ranging from 0 to 1.0. These values are multiplied for all land qualities for each land utilization type as follows;

LQ suitabilities = (LQ factor 1) x (LQ factor 2) x ....x (LQ factor n) ....(1)

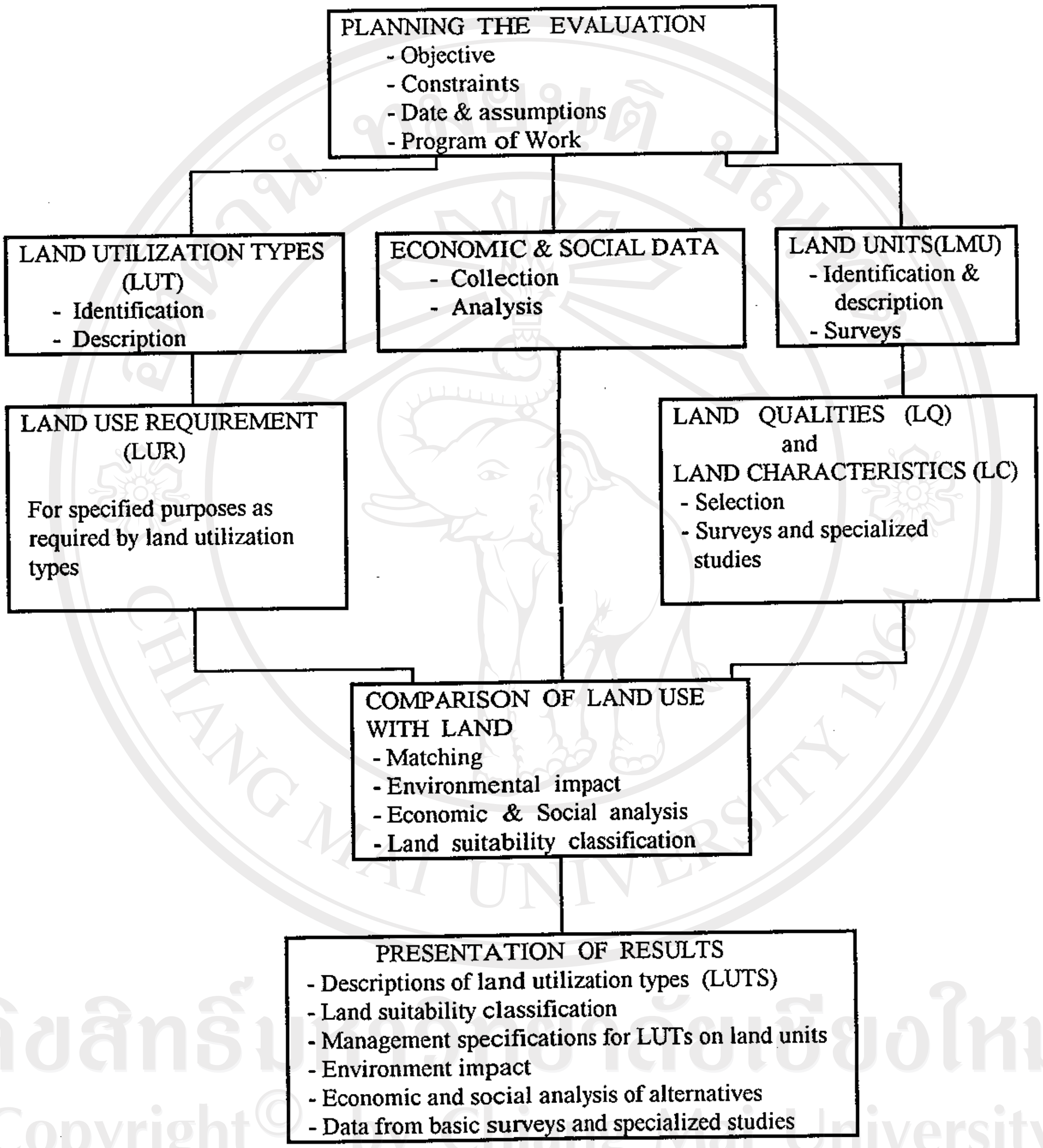

Figure 1. Procedures in land evaluation (FAO, 1984 and Young, 1986)

2.2.1.3 Modified Multiplication Method

Elbersen et al. (1988) modified the FAO multiplication method to classify suitability of land in Indonesia. Land quality was rated for each mapping unit. Minimum value of LQ was used as a weighting factor for overall LQ suitability by multiplying with the average value of LQ for the remaining factor as follow;

LQ suitabilities = min LQ factor **x** (average of remaining LQ factors/100) ......(2)

The results from the Multiplication method (equation (1)) and the Modified Multiplication method (equation (2)) were converted back to an overall suitability in accordance with the scale of crop yields related to guidelines for definitions of classes for factor rating. Factor rating for class S1 (Highly suitable) was assigned to expected crop yields of more than 80%, S2 (Moderately suitable) for expected crop yields between 40 - 80 %, S3 (Marginally suitable) for expected crop yields between 20 - 40 %, and N (Not suitable) for 0 - 20 %.

2.2.2 Automated Land Evaluation

The Automated Land Evaluation System (ALES) is one of widely used model (Rossiter, 1990). It is a computerized realization of the FAO's framework that allows land evaluators to incorporate their own knowledge base system with which they can compute the physical and economic suitability of land map unit (Fig. 2). Another system is LUPLAN, a program package which was developed to implement the SIRO-PLAN method of land-use planning devised by the CSIRO division of Land Use Research (Booth, 1986). However, these system was not developed to incorporate GIS capability.

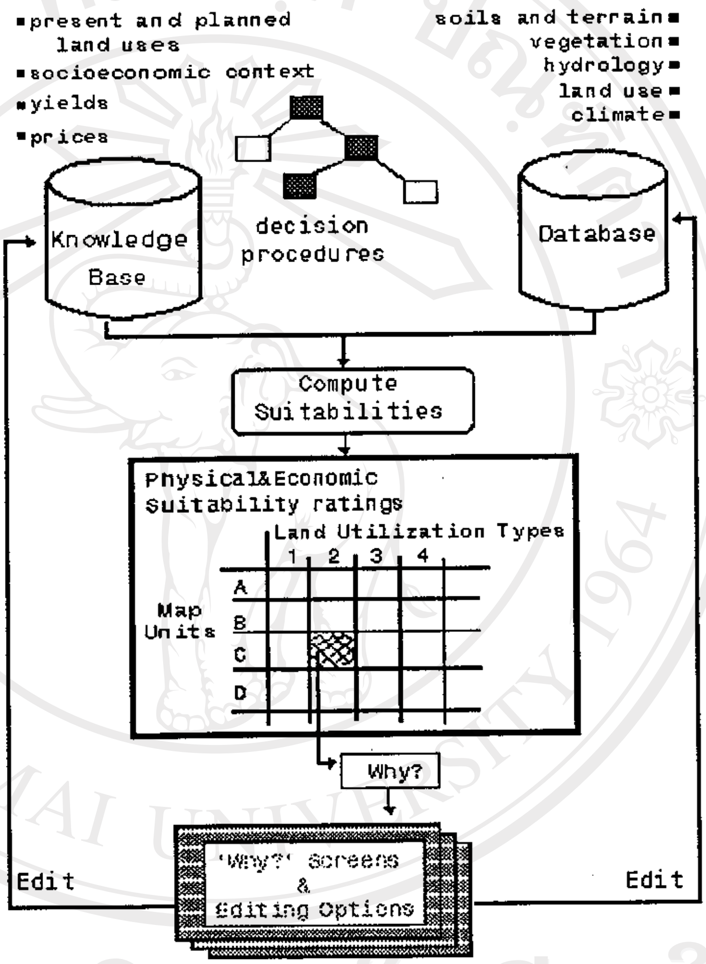

Figure 2. ALES program flow (Rossiter, 1990).

Many attempts have been made to automate land evaluation process by implementing on a microcomputer environment. The mathematical or quasi-mathematical models are used to describe the relations between the characteristics of land, water and climate and the desired land qualities. There are many kinds of models available ranging from a very simple to a highly complex. The kinds of models used in land evaluation are empirical models, deterministic process models and stochastic process models (Burrough, 1989a). Some examples of models for land evaluation are the Erosion Productivity Impact Calculator (EPIC) and the Productivity Index Model (PIXMOD). EPIC simulates potential productivity on the basis of the most important chemical and physical process that influence the growth of crops (Williams et al., 1983). PIXMOD was developed specifically for application in land evaluation studies in the wheat growing areas of western Canada (Onofrei and Smith, 1985).

2.2.3 Fuzzy Land Evaluation

GIS had been experimented with its overlay capability to produce land suitability maps. Since the linkage between the spatial entities and their non-spatial attributes are based on the membership concept of classical set theory. That is an entity either has an attribute entirely or does not have the attribute at all (Wang et al., 1990). Each spatial entity is associated with a single attribute. Sharp boundaries are imposed between the attributes. When assessing land suitability for a certain crop, a piece of land has to be categorized into either "suitable" or "unsuitable". In practical situations, there is not often a crisp boundary between suitable and unsuitable lands. Therefore, a fuzzy set (Zadeh, 1965) was introduced in land evaluation. Fuzziness is a type of imprecision characterizing classes that for various reasons cannot have, or do not have sharply defined boundaries. These inexactly defined classes are called fuzzy sets (Fig. 3). Fuzzy set theory provides useful concepts and tools for representing geographical information. In this case, partial rather than full or no membership allows for the analysis of information about more complex situations (Wang et al., 1990). Similarity Relation Model will be used to estimate the membership grades in this study.

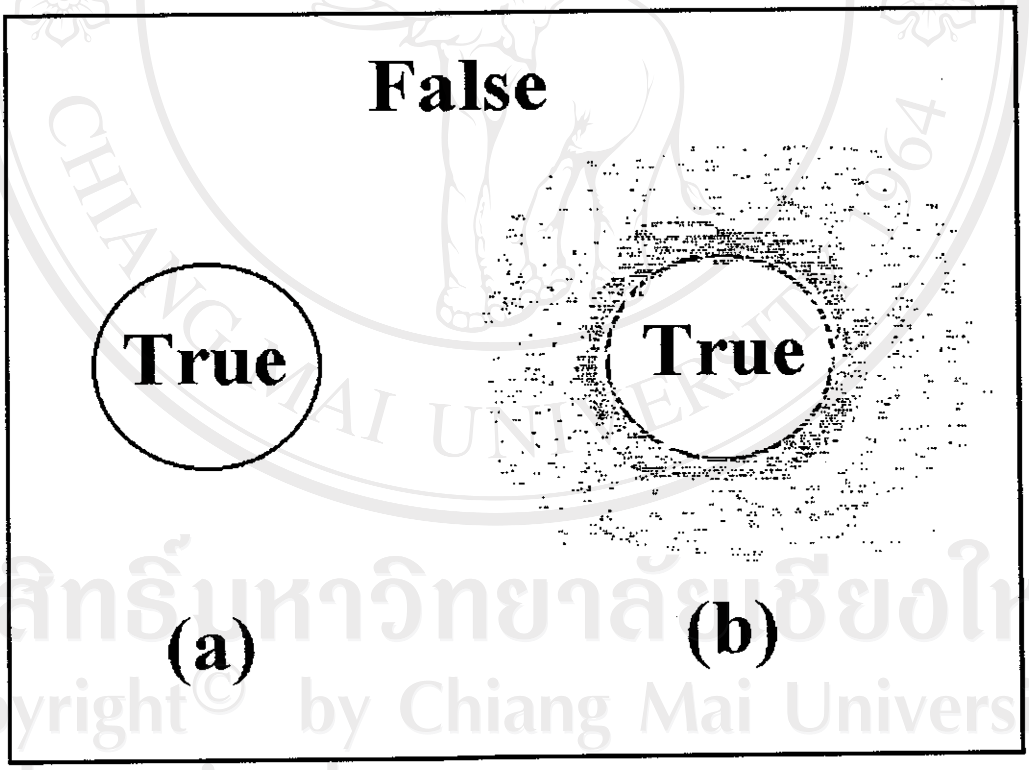

Figure 3. (a) Boolean set and (b) Fuzzy set (Burrough, 1989).

Wang et al. (1990) designed a fuzzy relational model to be used in GIS for assessing land suitability based on Similarity Relation Model. In this method, a GIS schema is defined. A schema consists of a definition of all application entity type including their attributes and relationships. Therefore, a schema identifies which data items should be stored and determines how best to organize them.

To assess land for its absolute suitability for a given crop, the rating of land suitability will be achieved by classifying areas into classes according to their land characteristics. By representing areas as vectors of a feature space, the distance between an area vector and a class will be used as a measure of their similarity. The similarity indicates the extent to which the area belongs to the class. The representative vector of class i can be shown in the form

$$(\mu_{i1}, \mu_{i2}, \ldots, \mu_{in})^{\tau} \qquad \ldots\ldots\ldots\ldots (3)$$

where; **n** is the number of physiographical characteristics to be considered and

μ_{ij} is value of the j^{th} characteristics of the requirements for suitability class i.

Let $\bar{x}$ be an area vector. The Euclidian distance between $\bar{x}$ and vector of class, $\bar{\mu}_c$

$$d_E(\bar{x}, \bar{\mu}_c) = \sqrt{\sum_{j=1}^{n} (x_j - \mu_{cj})^2} \qquad \ldots\ldots\ldots\ldots (4)$$

$$= \sqrt{(\bar{x} - \bar{\mu}_c)^{\tau} (\bar{x} - \bar{\mu}_c)} \qquad \ldots\ldots\ldots\ldots (5)$$

where; τ is the vector transposition.

The smaller the distance the more similar $\bar{x}$ is to **c** in terms of the land qualities.

An area can be associated with partial membership and belong to different classes to different extents. A fuzzy set is characterized by its membership function $(f_c(\bar{x}))$ Membership function (f_c) for suitability class **c** is;

$$f_c(\bar{x}) = \frac{\frac{1}{d_E(\bar{x},\bar{\mu}_c)}}{\sum_{i=1}^{m}\left(\frac{1}{d_E(\bar{x},\bar{\mu}_i)}\right)} \qquad \text{.................... (6)}$$

where m is the number of classes, and

$$\frac{1}{d_E(\bar{x},\bar{\mu}_c)}$$

serves as a normalizing factor.

For a given crop, four membership functions will be defined for the four suitability classes, S1 as highly suitable, S2 as moderately suitable, S3 as marginally suitable, and N as non suitable. By calculations, each area will have four membership grades indicating the extents to which this area belongs to each of the four classes.

To assess land for its relative suitability for different crops, the representative vector of the S1 of each crop will be assigned as the representative vector of that crop class. Wang defines the suitability of area $\bar{x}$ for crop **c** as;

$$s(\bar{x},c) = \frac{1}{\sqrt{(\bar{x}-\bar{\mu}_c)^{\tau}(\bar{x}-\bar{\mu}_c)}} \qquad \text{.................... (7)}$$

and the relative suitability for crop **i** as;

$$R_i(\bar{x}) = \frac{s(\bar{x},i)}{\sum_{j=c} s(\bar{x},j)} \qquad \text{.................... (8)}$$

R_i is actually the membership function of $\mathbf{f}_i$ and $R_i(\bar{x})$ is the membership grade of area $\bar{x}$ in $\mathbf{f}_i$. Each area can be associated with membership grades which indicate the relative suitability for the crops.

2.3 Map Outputs Comparison

The map similarity comparison has been investigated for over 25 years. Many studies have used the coefficient of correlation as the "expected" value for similarity. The classification error matrix typically contains tabulated results of accuracy evaluation for a thematic classification, such as a land-use and land-cover map. Diagonal elements of the matrix represent counts which are correct. The usual designation of classification accuracy has been total percent correct. Olson (1972a, 1972b) and Monmonier (1975, 1976) have looked at categorization to determine its impact on visual patterns. In a statistical study, the similarity of a pair of data sets was measured using Pearson's r correlation coefficient. The coefficients were compared before and after categorization, and the mean and standard deviation technique for creating categories best maintained the degree of correlation present in the original data. Carstensen (1984, 1986) studied the use of unclassed bivariate choroplethic maps for the determination of similarity. He showed that the unclassed bivariate map was suitable for this purpose. Lloyed and Steinke (1976) have provided useful findings concerning design and categorization for choropleth mapping. In many of these studies researchers used the correlation coefficient (Pearson's r) as a basis for comparison.

The KHAT statistics (also called "KAPPA"), proposed by Cohen (1960), describes the degree of agreement between two sets of categorical data.

KHAT was originally used in measuring agreement between opinions, but it has also been used in photo interpretation and remote sensing to measure agreement between classification schemes (Congalton and Mead, 1983). In remote sensing, the KHAT coefficient has been used to test different automated approaches to classification of multispectral data from the LANDSAT multispectral scanner (MSS) and thematic mapper (TM). KHAT has its greatest potential when incorporated into a GIS that utilizes a cellular data structure (Carstensen, 1987).

The value of the KHAT statistics range between -1 and +1 indicating complete disagreement or complete agreement. To calculate the KHAT statistics, an error matrix is developed. The columns of the matrix represent the categories on map number one, the rows are the same categories on map number two. Each sample point is represented in the matrix according to the categories found at its location on the two maps.

KHAT statistics is a comparison between two measures. The calculations for KHAT statistics can be done according to Congalton, 1991 as;

$$\mathbf{KHAT} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

where:

r = the number of rows in the matrix

x_{ii} = the number of observations in row i and column i

x_{i+} = the marginal totals of row i

x_{+i} = the marginal totals of column i

N = total number of observations

In summary, land evaluation can be achieved by many methods each has different algorithm which require different resources and yield different forms of outputs. This thesis attempts to use fuzzy method to evaluate land in Prao district for their suitability and develop a computer program which link land evaluation system with GIS to show the results as maps. It also attempts to develop the "Shell" to help user in using land evaluation system with minimum efforts in computer work.