LAND SUITABILITY EVALUATION: IMPROVING ACCURACY OF ASSESSMENTS WITH A NEW PARADIGM BASED ON GEOSTATISTICAL ESTIMATION AND FUZZY SET THEORY

A Thesis Submitted to the Committee on Graduate Studies

in Partial Fulfillment of the Requirements for the

Degree of Master of Science

in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Watershed Ecosystems M.Sc. Program

June 2000

0-612-57975-1

To my Dad who passed away eagerly waiting
to see this work reach completion
may he rest in peace

ABSTRACT

Land Suitability Evaluation: Improving Accuracy of Assessments with a New Paradigm based on Geostatistical Estimation and Fuzzy Set Theory

Hamad Omar Mohamed

Conventional methods of soil classification are map unit based, consisting of discrete, sharply bounded areas. In this study an alternative approach to land suitability evaluation based on conventional methods is suggested to account for variability within soil classes and mapping units. This approach combines optimal spatial interpolation methods (Block Kriging) and Fuzzy Set Theory to derive a continuous classification and mapping for land suitability evaluation by retaining point-data from field. This proposed paradigm was applied to the land suitability assessment for maize in the Texcoco River Watershed in central Mexico. Ten polygons mapped by conventional survey procedures represent the "current paradigm" of generalized soil information whereas 66 point-data represent the proposed "new paradigm". The soil data from these paradigms were processed and evaluated by an Automated Land Evaluation System (ALES) model. Suitability maps were generated from both methods. The final suitability maps were compared in terms of RMS with observed performance (maize corn yields), in order to test their accuracy. Land suitability evaluation results from interpolation and fuzzy classification methods showed significantly closer predictions to observed yield than the discrete (polygon) classification. There was no significant difference, in terms of accuracy of predictions, between estimates from Kriging interpolation alone against estimates derived from fuzzy membership application. There are advantages, in terms of accuracy of interpretations derived from soil data, in retaining hard point-data and then using spatial interpolation

and fuzzy membership functions to derive interpretive maps. Geostatistical techniques and Fuzzy Set Theory and algorithms, used in a Geographical Information System (GIS), are promising tools to avoid information losses due to generalization.

ACKNOWLEDGEMENTS

I am grateful to Allah for all the favors He has bestowed upon me. Special thanks go to my supervisor, Dr. Raul Ponce-Hernandez, for his unenviable task of supervising this thesis and for his great effort in providing me with all data and information needed to complete this work. I wish to thank all people in the Watershed Ecosystems Department who helped in making this work a possibility. I am thankful to my committee members, Dr. Tom Hutchinson and Dr. Jim Buttle, for their helpful and constructive comments. I would like to thank Eric Sagar, Robert Loney and Paul Karrel for their directions in laboratory and computer work which was most helpful. Many thanks and great appreciation go to my wife Madiha for all her endurance and tremendous support throughout the development of this thesis, and to my lovely son Mohamed and daughters, Esra and Fatma, for their patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF MAPS

1 INTRODUCTION

The ability of land to produce crops is limited and the limits to production are set by climate, soil conditions, the genetic potential of the crops and by the use and management of land.

The assessment of land potential is an essential component of land use planning and requires a comprehensive exercise in land evaluation. Such an exercise involves detailed data on the physical, chemical and morphological characteristics of land resources, land uses and the technical, infrastructural and economic settings surrounding such land uses. Accordingly, knowledge of the land resources endowment and its potential is an essential prerequisite to planning optimal land use and subsequent sound "long term" agricultural development.

The term "land" has been used in a comprehensive integrating sense to refer to a wide variety of natural resource attributes, ranging from the near atmosphere down to the sub-soil and the underlying rock. Land evaluation is the process of assessing the potential uses of land for agriculture, engineering, forestry and recreation. Specific agricultural uses include arable farming, extensive grazing and irrigated agriculture. Land evaluation is based on the interpretation of physical land attributes with respect to kinds of land use and the extent to which crop production or the given target performance of the land-use, can be achieved optimally and on a sustained basis (i.e. without deteriorating of the land resources). The relevance of the land use for the economic and social context of the area concerned should be also accounted for. Additionally, land evaluation determines the best management and improvement measures for each alternative kind of use.

Land use decisions are based on the characteristics of the land which need to be inventoried and mapped. In turn, the characteristics of the land may show considerable variability over space and over time. Thus, the information and the accuracy of the statements that can be made about soil properties depend to a large extent on the spatial variability of the soil and on the accuracy with which it is characterized and portrayed in maps. The effects of inaccuracies introduced by mapping methods and techniques on the truthfulness of the statements that can be made about soil variability and on the predicted values of soil properties, and subsequently on the suitability ratings assigned during land evaluation, are unknown and need to be investigated.

Land suitability classification is an approach in land evaluation that concerns the appraisal and grouping of specific areas of land in terms of their suitability for specific uses (FAO, 1976). The Food and Agricultural Organization of The United Nations (FAO) proposed general classification for lands suitability is universally accepted in land-use planning, particularly in the developing world. In this method, the suitability classes are defined as discrete groupings, separated by strict class definitions or fixed class limits. Land units that have a degree of suitability somewhat intermediate between classes can, however, only be placed in one single suitability class (Tang et al., 1991).

Thus, the problem of characterizing and mapping of soil spatial variability through the use of discrete classification systems in land evaluation needs to be examined thoroughly for its implication for land-use planning. Many conventional soil classification systems establish a series of subdivisions, which place individual soil profile descriptions into a hierarchically structured, and to a great extent rigid, scheme. Examples of such a scheme include the United States Department of Agriculture

(USDA)'s comprehensive "Soil Taxonomy" (Soil Survey Staff, 1975), Northcote's (1979)'A factual key to Australian soils', and FAO's Soil Classification System (FAO, 1989). These hierarchically arranged classes are mutually exclusive with sharp boundaries between class limits. Each class identified is defined by a central concept (Burrough,1989). The conventional method outlined above therefore implies that soil classes are discrete with abrupt boundaries, represented by a central concept and makes no allowances for either class impurities due to missing or unrecorded data or vagueness in the definitions of the class boundaries. According to the existing systems of soil classification, any soil individual belongs to exactly one class.

This study focuses on comparing the effects of applying the current and the "proposed" paradigms for spatial soil/land resources representation. Particular attention is paid to the accuracy and practical value of interpretive maps for land-use planning such as land suitability maps derived from them.

Based on measurements from the Texcoco watershed in Central Mexico, the objectives of this study were the following:

1. to test the validity of the current paradigm of soil variability representation against a set of methods, procedures and models, which make up a "proposed paradigm"; 2.to compare (quantitatively) the virtues of both current and "proposed" paradigms of soil spatial variability representation in terms of accuracy of predictions and interpretations; and 3.to assess the practical implications of using each paradigm for soil/land suitability interpretations, particularly in terms of errors and in terms of applicability.

The proposed methodology aims at examining the practical implications of the paradigm shift from the conventional mapping and interpretation of soil/land information, as used in current land evaluation and suitability interpretations, to the proposed paradigm that utilizes a different way of dealing with spatial variability and spatial interpolation and with uncertainty in mapping. The methods designed will compare the current and the "proposed" paradigms (proposed by this research) in terms of the accuracy of interpretative classifications after land suitability evaluation using expert system models (Automated Land Evaluation) and raster maps derived from spatial analysis and interpolation of point data by Kriging and other interpolators.

Under the proposed paradigm, boundaries of such maps will recognize the uncertainty in class allocation and in mapping boundaries by representing transitions by means of membership functions in Fuzzy Sets. This membership function has values between 0 and 1. Unlike Boolean sets, fuzzy sets can overlap and an individual can be a member of the overlapping sets to different degrees (Burrough and McDonnell, 1998). Predicted grain maize yields derived by both current and proposed methods (including interpolated and fuzzified methods) will be compared against observed maize yields. Finally, the different methods applied to determine land suitability will be discussed in terms of their efficiency and accuracy for representing and mapping soil spatial variability.

2 BACKGROUND

2.1 Current Paradigm for Providing and Processing Soil Information for Land Use Planning

Soil survey and soil classification have traditionally been the most practical approach to grouping similar and separating different soils on a regional scale. However, soil survey and classification, in the traditional sense, are based on the notion that soil forms discrete, internally uniform units, with sharp boundaries at their edges (Odeh et al., 1992). Therefore, the conventional methods imply that soil classes are discrete and discontinuous with sharp boundaries, and that they can be represented by a central concept known as a typical profile (Mazaheri et al., 1995). The ultimate use of soil profile classification is to establish a map with a set of clearly defined and mutually exclusive classes that can be used for transferring information about the soil. The soil map produced by conventional concepts is a single display of the spatial distribution of the classes of the initially constructed classification scheme, with soil boundaries interpolated between points where the soil is allocated to different classes (Burgess and Webster, 1984). Information regarding soil properties over a given area is usually derived from the existing soil survey maps in which the carefully drawn lines give the impression that the map unit (usually represented as a polygon) is relatively homogeneous with sharp boundaries (Burrough, 1983).

2.1.1 Conventional Soil Survey

This section will focus on explaining definitions and concepts of soil survey. The key factor in determining what kind of soil survey needs to be taken and what properties are to be measured and/or observed is the ultimate use of the survey. The usefulness of soil survey depends on two things: the accuracy with which soil properties are mapped out, and the relevance of those properties to the purpose at hand (Dent and Young, 1981).

A soil survey is very important for soil management because it is the source of soil information which is needed by land use planners to make decisions about the suitability of land for a variety of purposes. For land evaluation, the basic questions are: Given a soil survey, what use can we make of it? How reliable is it? How specific are its statements? To answer these questions we must understand the purposes and kinds of soil survey, and how they are made.

2.1.1.1 Field Procedures

The steps in soil survey are soil description, classification, analysis and soil mapping. These will determine the types and pattern of occurrence of soil mapping units (polygons on a map) on the landscape and the drawing of them on maps.

Avery (1978, p.1) gives the most general definition of the objective of soil survey: "The general aim of soil survey is to provide information about the soil of areas of land". This definition includes any systematic soil investigation, not just mapping, and maps of any type (classes, single-factor, etc.). Nevertheless, most authors consider mapping a fundamental part of soil survey (e.g., Eyk et al.,1969). The primary purpose of a soil survey is to recognize and identify three-dimensional bodies of soil which have significance for some particular objective, and to plot their geographic distribution on a base map. Dent and Young (1981) emphasized the same idea, which is to concentrate on the main objective or purposes of soil survey so as to define the type of soil survey needed. The practical purpose of soil survey is to enable numerous, more accurate and more useful predictions to be made for specific purposes than could have been made otherwise (i.e., in the absence of location-specific information about soils). Again the USDA (1982) definition emphasizes the objective of soil survey. A soil survey describes the characteristics of the soil in a given area, classifies the soil according to a standard system of classification, plots the boundaries of the soil on a map, and makes predictions about the behaviour of soils.

The different uses of soils and how management affects them are considered in designing and carrying out the survey. The information collected in a soil survey helps in developing land-use plans and in evaluating and predicting the effects of land use on the environment.

From the previous definitions it appears that the main objectives and purposes of soil survey include the following:

- Establish and research the relationships that exist between soil morphology and other soil properties of interest.

- Provide information and describe the characteristics of soil which can be interpreted for a wide variety of land use purposes.

-Classify the soil according to a standard system of classification.

-Produce soil maps with associated reports.

-Make predictions about the behaviour of soils.

-Help the farmers, foresters, engineers, planners, development agencies and other users to make wise decisions about land use and land management.

Even though the main reason for soil survey is to show the geographic distribution of soils, there are important differences in the objectives of surveys. Thus, a soil survey may be one of two kinds: general purpose or special purpose soil survey.

General-purpose soil surveys are expected to provide the basis of interpretations for many different purposes, some of which may not yet be known. General purpose soil survey involves the production of a pedological map, which shows the distribution of soil units defined primarily according to their morphology, and the acquisition of field and laboratory data on other physical, chemical and biological characteristics of these units. These surveys can be used for many purposes and are most useful where little is known about the soil cover, particularly in less developed regions. However, these surveys cannot be used for specific purposes. Some data collected by this type of survey may never be used due to their generality. **Special-purpose soil surveys** are carried out for a specific and known purpose (e.g. an irrigation project or any other conservation- oriented farm planning). The advantage of this type of survey is that it is very rapid because the purpose is well known so the surveyor can concentrate on properties of interest. Generally speaking this type of survey is less expensive than general-purpose surveys.

In the field the soil can be sampled by either a **free survey** or a **grid survey**. In the free survey the surveyor is free to choose sample sites according to a prior study of the climate, geology, geomorphology, vegetation, land use and land use history of the area. In this case the surveyor will predict the relationships in the landscape and use aerial photographs to draw the boundaries between different mapping units. The surveyor's judgement and experience are very important. The free survey is more efficient than grid survey because its sampling strategy is based on aerial photograph interpretation. In grid survey statistical methods are used to divide the space into a regular rectangular grid over the survey area. This is useful for large -scale intensive survey to take into account the spatial variability and investigate the relationships between properties in a complex area or where there are no morphologic controls on soil properties. A major disadvantage of a grid survey is that it is inherently wasteful; a significant proportion of sites are unrepresentative, including, for example, settlements, or near landscape boundaries when the soil class is indeterminate. Inflexibility of site selection can also be a severe disadvantage where access is frequently interrupted by creeks, dykes and so on, so that time is wasted reaching the specified sites (Dent and Young, 1981). What seems to be a useful approach to overcome the disadvantages of both free and grid survey is to combine the two types of survey. This allows the surveyor to cover most of the area, and at the same time enabling him or her to choose other sites for observations according to his or her judgement.

2.1.1.2 Soil Classification

Soil classification is the categorization of soils into groups at varying levels of generalization according to their morphological properties and/or assumed genesis important for the objectives of the classification (Buol et al., 1989). The purpose of any classification is to organize our knowledge so that the properties of objects may be remembered and their relationships may be understood most easily for a specific objective. The processes involve formation of classes by grouping the objects on the basis of their common properties. Any system of classification groups soils so that a greater number of most precise and most important statements can be made for the objectives of the survey (Rossiter, 1994). Soil classification plays an important role in predicting soil properties at unknown places and transferring soil management technology from one place to another.

There are two kinds of classification (i.e. technical and natural classification) and we differentiate between them as follows:

Technical classification: organization of objects in a grouping (either single level or multicategoric) for a specific applied purpose.

Natural or scientific classification: categorization in which the purpose is to bring out relationships of the most important properties of the population being classified, without reference to any single specified applied or utilitarian purpose (Cline, 1949).

Only two systems of soil classification enjoy very wide international recognition: Soil Taxonomy (USDA, 1982) and the Soil Map of the World (FAO-UNESCO, 1974). The Soil Taxonomy is a hierarchical system with six levels of detail, each contained in the next-highest category. Each class has a central concept and a range of variation of the properties defined. The class represented by the central concept has rigid limits. On the other hand, the FAO-UNESCO Soil Map of the World is the source of soil data in many areas of the world. The objectives of the Soil Map of the World (FAO, 1974) are to:

- make the first appraisal of the world's soil resources;

-supply a scientific basis for the transfer of experience between areas with similar environments;

-promote the establishment of a generally accepted soil classification and nomenclature;

-establish a common framework for more detailed investigation in developing areas;

-serve as a basis document for educational, research, and development activities; and

-strengthen international contacts in the field of soil science.

A decision on the soil classification is usually made in advance of fieldwork. If the USDA system is used, there is a commitment to a considerable amount of analysis necessary to establish the classification of each soil unit. The classes in the legend can be locally defined, or taken from external systematic classification. In the local classification, classes are established locally on observed differences between soils as they occur in the field. They are generally given local names. In the systematic classification the classes are established in some hierarchical taxonomical system, and the local soils must fit into one of the existing classes.

The decision on choosing one of the classification systems depends on the purpose of the classification and the level of the generalization attempted. If the purpose of the survey is precisely defined, then the soil properties are to be measured at specific known sites. Then the soils from these sites can be grouped on an ad hoc basis. In contrast, the general purpose or taxonomic classification usually is of national scope and only general predictions of soil properties and their response to management can be made. In the field, the surveyor examines soil profiles, which are point samples of the soil cover. These sample descriptions are then stored into conceptual groups, or taxonomic units. Each of these are defined by a typical profile form (that is, the central concept), and an allowed range of variation around it (Dent and Young, 1981).

One of the major problems that arises with the use of discrete classes in soil mapping is that classes are discrete. Therefore sharp cut-offs have to be imposed in the character space that disregard the continuity of the soil. Furthermore, if these are projected onto the geographic space the continuity here is also lost (McBratney et al., 1992). Another assumption made in soil classification systems and in soil survey practice is that soil differences can be adequately characterized by relatively few diagnostic properties that are used to define categories. Taxa of soil classification systems are ideally those which have greatest independence of variation from each other but which have high covariance with many other nondiagnostic properties. This results in the variance within taxonomic units measured over all properties being minimized with respect to their total variance (Trangmar et al., 1985).

2.1.1.3 Soil Mapping

Soil mapping first involves classification. The two processes cannot be separated. The surveyor maps out the soil continuum as a pattern of soil areas. Each mapping unit is supposed to have the same properties throughout and the selected profile is supposed to represent the variability of these properties. Almost all boundaries may be drawn on the basis of aerial photograph interpretation and field observations. Mapping units are not the same as taxonomic units. Mapping units are real soil areas. The surveyor inserts the boundaries on the soil map which depict as faithfully as possible the perceived situation and which delineate units that are as homogenous as possible. The decision as to whether to use simple mapping units, i.e. those containing soil of only one taxonomic unit, or whether to use compound mapping units which contain soil of two or more taxonomic groups, needs to be made. The decision depends on the complexity of the soil pattern and the purpose of the soil survey.

In reconnaissance survey almost all boundaries are drawn during photo intepretation. The surveyor will concentrate on the transitional zone to make a clear separation between two classes. However, if the observations (auger holes, pits) are spaced closely enough we can draw the boundaries by eye. Nevertheless, there is a real problem with false precision. The other option is using the landscape analysis approach in which the boundaries we draw are visible over their whole length on the landscape.

The scale of the published map sets limits to the amount of detail that can be shown and the level of generalization. As published map scales become smaller, so the minimum size of a particular mapping unit that can be shown increases.

The results of a soil survey are presented in a map and associated report. The ease with which the map can be used depends upon the cartographic quality and the clarity of information on the legend. A more general legend makes fewer separations, based on few categories and more general classes than a specific legend.

Generally, the soil map, or indeed any type of map, is used as a predictive tool; the aim is to indicate the nature and properties of soil at one or more points or areas. A soil map is designed to answer any specific types of questions and any user of the results should be aware of the associated errors. Subdivision of a landscape into units that are separated from each other by sharp boundaries is unnatural, because natural boundaries tend to be gradual rather than abrupt (Bouma, 1989). Ideally, map units should be 100% pure, as assessed by a set of randomly selected sites or check points, which are used to find the dominant soil class. In practice, the purity of mapping units does not exceeds 55% and in the best of cases is around 65-75% (Beckett, 1984). In their review of this topic, Beckett and Webster (1971) concluded that simple mapping units might actually average only 50% purity. Burrough et al. (1971) found that mapping purity varied with map scale and observation density. Purity ranged from 45-63% at a scale of 1:63,360 to 65-87% at a scale of 1:25,000. This apparently high degree of taxonomic variability within mapping units is often underestimated in its importance because impurities often differ only in minor definitive features and do not require different management (Bascomb and Javis, 1976). A further problem that arises in a field survey is that soil units which are practical or easy to map are not necessarily equivalent to the taxonomic classes of a standard system of soil classification.

Finally, it must be noted that thematic mapping, which can be considered as a special purpose survey, can produce maps at large scales by computer automation for every soil property and for any plant nutrient. Such a procedure has a great potential for use in land evaluation and the greatest predictive value for land characteristics if knowledge of spatial variability is incorporated into the spatial-prediction algorithm (Ponce-Hernandez,1995).

2.1.2 Interpretation of Soil Information

It is well known that soil survey is very important in assessing and aiding in land use and management planning from scales ranging from the global to the individual farm. Soil survey produces a range of information required for its interpretation and application, including land use potential, management practices, plus data needed to be used as a basis for economic evaluation. The soil survey users are interested in interpretations, i.e. what the survey says about actual and potential land uses and land management strategies. If we consider soil as a part of the land resource, soil survey has become an input to land evaluation. The basic idea is to interpret detailed information from soil survey reports in terms of land suitability and land capability classifications.

2.1.2.1 Land Capability Classification

It is relevant to define some important concepts that are involved in the process of land evaluation. **Land** - according to FAO (1976) land is an area of the earth's surface, the characteristics of which embrace all reasonably stable, or predictably cyclic, attributes of the biosphere vertically above and below this area. **A land unit** is an area of land distinctively different in its attributes from others, and which possesses sufficient internal uniformity in its characteristics relevant to its management that can be managed in the same way. A land unit can be treated as a single entity, and is the result of inventory and surveys. A land unit can become a land management unit (Ponce-Hernandez, 1995). **Land evaluation** is the process of estimating the potential of land for alternative kinds of use. These include productive uses, such as arable farming, livestock production and forestry. These are uses that provide services or other benefits, such as water catchment areas, recreation, tourism and wildlife conservation (Dent and Young, 1981). Land evaluation involves a comparison between the requirements of the land use and the qualities of the land (Dent and Young, 1981). **Land Capability** is the potential of the land for use in specified ways, or with specified management practices (Dent and Young, 1981). **Land Limitations** are land characteristics which have an adverse effect on capability. The assessment of land capability involves an evaluation of the degree of limitation posed by permanent or semi-permanent attributes of land to one or more land uses. It is essentially a negative approach whereby as the degree of constraint increases, so land is allocated to lower classes. Land capability assessment is based on a broader range of characteristics than soil properties. Information on slope angle, climate, flood and erosion risk, as well as on soil properties is required. The main product of land capability classification is a map in which areas of land are put into capability classes ranging from best to worst. The prime objective of the method is to assess the degree of limitation to land use potential imposed by land characteristics on the basis of permanent properties. A scale of land capability can thus be envisaged with the degree of limitation and hazard defining the class. As the degree of limitation increases, so the range of land use option decreases. There are three categories recognized: capability classes, subclasses and units. Capability classes are groups of land units that have the same degree of limitation. The risks of soil damage or limitation become progressively greater from class I to class VII. The USDA capability classification is one of a number of the interpretative groupings based primarily on interpreting major kinds of land use. The classes show the general capability of a land unit for agricultural use. Capability subclasses are defined on the basis of major conservation problems, such as:

e- erosion and runoff

w- excess water

s- root-zone limitation

c- climatic limitation.

The capability subclass provides information on the kind of conservation problem or limitation involved. Class and subclass together provide the map user with information about both the kind of problem involved and the degree of this limitation. A capability unit is a subdivision unit of class on the basis of potential productivity. All soils within a sub-class having comparable potential productivity belong to the same capability.

This means that soils in a capability unit are sufficiently uniform to:

a) produce similar kinds of cultivated crops and pasture plants under similar management practices;

b) require similar conservation treatment and management; and

c) have comparable potential productivity.

2.1.2.2 Land Suitability: The FAO Land Evaluation Framework

Land suitability is the fitness of a given type of land for a defined use (FAO 1976). Suitability is a statement of the adaptability of a given area for a specific land use. By narrowing down the range of land uses considered, it is possible to be more specific about the fitness of the land for a given use, this being implied by the word suitability rather than capability.

The FAO proposed general classification for land suitability is universally accepted for land use planning purposes, particularly in the developing world. In this system two suitability orders are discerned:

suitable (S) and unsuitable (N). The order S is subdivided into a very suitable

(S1), moderately suitable (S2) and marginally suitable (S3). The order (N) is subdivided into currently unsuitable (N1) and permanently unsuitable (N2). A suitability map shows the suitability of each land-mapping unit for each defined kind of land use.

The objective of land evaluation is to judge the anticipated performance of an area for defined purposes. The framework (FAO 1976) is designed so that through land evaluation a user should be able to answer questions of the following type:

How is the land currently managed, and what will happen if present practices remain unchanged? What improvements in management practices, within the present use, are possible? What other uses of land are physically possible and economically and socially relevant? Which of these uses offer possibilities of sustained production or other benefits? What adverse effects, physical, economic or social, are associated with each use? What recurrent inputs are necessary to bring about the desired production and minimize the adverse effects? What are the benefits of each form of land use?

The evaluation process does not in itself determine the land use changes that are to be carried out, but provides data on the basis of which such decisions can be taken. Various steps are necessary in order for the evaluation exercise to answer these types of questions. In the first instance there must be a clear statement on the objectives of the study.

Selection of relevant land characteristics (attributes of land which can be measured or estimated) is possible only within the context of a particular study. Two strategies are possible according to the FAO framework once the objectives of a study are stated **(Fig.1)**. In the two-stage approach, an economic and social analysis may follow from a qualitative land classification, while in the parallel approach the analysis of the land and land use relationships proceeds concurrently with economic and social analysis.

The parallel approach is expected to give more accurate results in a shorter period of time. It offers a better chance of concentrating survey and data collection activities on producing information needed for the evaluation. The suitability evaluation involves relating mapping units to specified types of land use. A distinction is made between **a major kind of land use** and **land utilization type.** The former is a major division of rural land use (for example, pasture land, forestry or recreation), whilst the latter is a type of land use described in greater detail (FAO, 1976).

It is also possible to have multiple land utilization types, and this term refers to a situation in which more than one kind of land use is practiced within an area. A land quality is a complex attribute of land which acts in a distinct manner in its influence on the suitability of land for a specific kind of use. It is assessed from land characteristics which are attributes of the land that can be measured or estimated (for example, slope angle, rainfall, and soil texture). Thus, the framework for land evaluation recommends that the land should be evaluated for land utilization types in terms of land units. To achieve such an evaluation. diagnostic criteria are recognized: these may be land qualities or characteristics, but they are known to have a clear effect on land use output or potential. Critical values are associated with diagnostic criteria so that suitability classes can be defined.

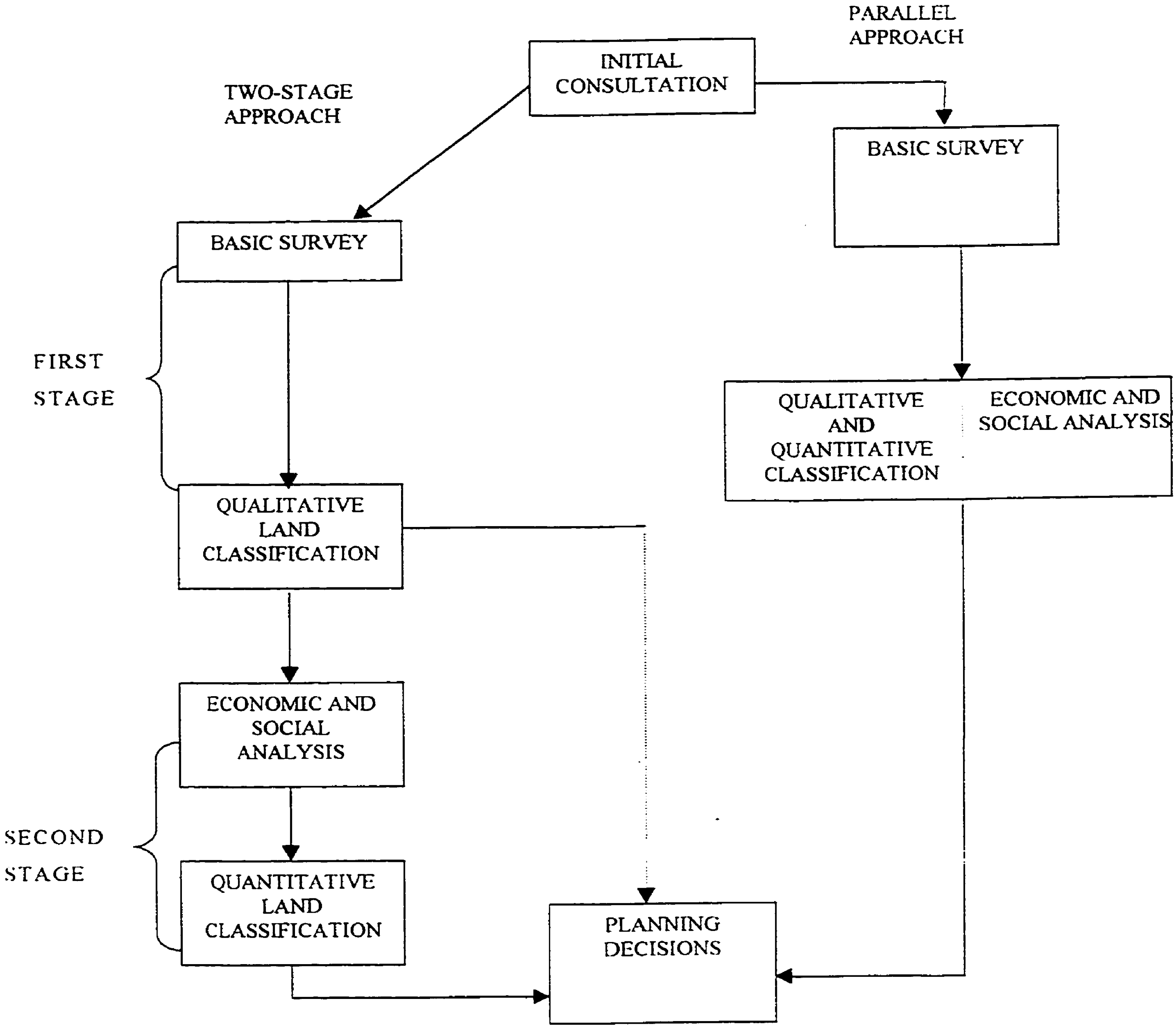


Fig.1: Two-stage and parallel approach to land evaluation (FAO, 1976)

2.1.2.3 Interpretation in Terms of Agro-Ecological Zones (AEZ) Approach

The term Agro-ecological Zones (AEZ) refers to areas of land which have been delineated by a method of dividing the earth's surface into relatively homogeneous areas with respect to the physical factors that are most important to crop (or plant) production. The term first came into prominence with the FAO's effort of the mid 1970s to determine potential human carrying capacity (Davidson, 1992).

The starting point of the procedure is the preparation of land resources databases in the form of digitized maps to define several components for each mapping unit. In the Agro-ecological Zones project, inventories of crops were prepared based on their climatic requirements (including moisture, temperature, radiation and photoperiod), and their effect on crop growth and phenology. The combination of available water and adequate temperature for crop growth is expressed in the growing period. The soil requirements include internal requirements (e.g. soil temperature, moisture, aeration, fertility, depth, stoniness, salinity and others) and external requirements, such as slope and flood conditions. Potential yield is then calculated for major crops, and the result is predicted yield as a percentage of potential yield for each crop. Ponce-Hernandez (1998) developed a methodology for ecological and economic zoning for the Amazon Basin based on the AEZ principles. The main aim for the project was to provide flexible methodological guidelines and procedures for zoning the Amazon basin based on both ecological and economic criteria. The methodology consists of five main stages:

a) The compilation and development of spatial and attribute databases of natural resources (Bio-physical databases);

b) the integration of the thematic maps and their attributes from (a) above (e.g. soil units, agro-climatic units, vegetation, etc.) into Ecological Zones or Ecozones;

c) the compilation and development of land-use spatial and attribute databases;

d) the integration of land-use variables into Land Utilization Types (LUT); and

e) evaluation or assessment of suitability of Ecozones for LUT.

Typically, land evaluation is an exercise which demands fairly large volumes of data and therefore a considerably long processing time, if done by hand. Such volumes of data and processing times demand automation, which is becoming a standard practice in land suitability assessment exercises. For instance, **ALES (Automated Land Evaluation System)** is a computer program that allows performance of land evaluation processes (Rossiter and van Wambeke, 1995).

The principles of the FAO framework were used by Rossiter and van Wambeke (1995) to develop the automated Land Evaluation System software that allows land evaluators to build models that can be used to compute suitability assessments. Each evaluation model consists of a set of proposed land utilization types, a set of outputs, and a set of land characteristics. Each land utilization type is specified in terms of its land use requirements and outputs. Each land use requirement within a land utilization type has a model-builder-specified number of severity levels (i.e. levels of limitation of the corresponding land quality). The model builder describes to ALES how to determine the severity of each limitation on the basis of land characteristics.

Land characteristics have a format defined by the model builder, thus allowing the use of any data in any form. Usually they are discrete, and have a user-defined number of classes, each with its own code and continuous values. These are related to commensurate discrete characteristics for further use in the knowledge base. The model builder also constructs data entry templates which control how data can be entered by the model user, or be read from other data bases. The way in which the ALES model builder reasons with classified data is reflected in the model in the form of decision-trees. Decision trees are hierarchical multiway keys in which the leaves are results such as land quality ratings, and the interior nodes (branch points) of the tree are decision criteria such as land characteristics values. These trees are constructed by the model builder, and traversed during the computation of an evaluation result, using the actual land data. The result is a complete land suitability classification for both physical and economic mapping units (Rossiter and van Wambeke, 1995; Ponce-Hernandez, 1989).

3.0 STATEMENT OF THE PROBLEM

Most natural properties of the earth's mantle vary continuously. However, the observations that describe these properties which form the databases of GIS are usually fragmentary. This is because, in general, we can observe at only a finite number of the infinity of possible locations. Even where a complete cover of information exists, for instance from satellite images, we often need to sample because the resolution of the data is too coarse for the purpose at hand, or there are too many data to handle or analyse in any reasonable time (Atkinson et al., 1990). Even for a small area and over short distances soil properties show great spatial heterogeneity. As much as half the variance of soil properties present within 1 ha is already present within a few metres (Becket and Webster, 1971). Soil properties vary also with depth. Different treatments affect the soil to different depths (Askew and Rigg, 1932). Nutrient or water uptake is not always from the same depth in different soils. So, soil variability is not the same at all depths, nor does it change with depth in the same way in all seasons or for all properties (Beckett and Webster, 1971; Raupach, 1951; Towner, 1968).

Soil survey maps are a source of very important quantitative and qualitative information, which can be used for land evaluation. The traditional model of discontinuous soil variation represented by choropleth mapping (Hole and Campell, 1985) implies that the predicted value of a soil property at any unsampled location within the mapping unit is the value for the typical pedon or the mean value for the mapping unit. Regrettably, this has led to the belief among many soil-map users (and indeed users of maps for other natural resources) that delineated units on maps are internally uniform with respect to specific diagnostic characteristics (Lyford, 1974). However, many studies (e.g. Becket and Webster, 1971; Burrough, 1987) have shown that within-unit variance is often unacceptably high for soil maps produced by conventional methods. Conventionally, soil and landscape classification proceed by identifying the central concept of class. Thereafter, the class limits are defined, usually in terms of a set of discriminating criteria. Most commonly, the boundary values delineating the class are sharply defined; for example, in Soil Taxonomy (Soil Survey Staff, 1976) the lower limit of organic matter for a histic epipedon is set at 14%. In this type of classification model it is implicitly assumed that all change between classes takes place at the class boundaries and that within the classes little change of importance occurs. The conventional model of spatial classification divides a landscape into the units of a choropleth map in essentially the same way. The basic entities (the data models of the phenomenon of interest) are then so-called homogeneous mapping units separated by sharp boundaries. This method is used even though spatial changes in the field may occur gradually over distances that are substantially greater than the narrow zone covered by the thin line drawn on the paper (Burrough, 1998). In this case the classes are discrete and therefore sharp cut-offs have to be imposed in the character space disregarding the continuity. Furthermore, if these are projected onto the geographic space, the continuity here is also lost. This has been a major problem in the preparation and use of soil (class) maps (McBratney et al., 1992).

The central purpose of soil classification and soil mapping, as products of conventional soil survey, is to enable a user to predict values of any given soil property at any specific depth and at any specific geographical location within a mapped area, without having to go there to observe or measure again. Soil classes and soil mapping units are the main sources of data for practical applications, such as in land evaluation. Therefore, these products are central to procedures for land evaluation, site suitability assessments and many other practical applications that require data on soil and land. Many conventional soil classes and their mapped spatial extents in conventional (choropleth) soil maps are being converted from analogue to digital form via digitizing, during the development of spatial databases in many organizations of many countries. In turn, these databases are utilized as sources of information for land evaluation and decision -making regarding land use. However, there may be fundamental problems with the information derived from the conventional soil classification and soil maps.

The validity of the interpretations, for practical purposes, that are derived from such information, whether aided by models and/or by analytical procedures in Geographical Information Systems (GIS), needs to be assessed and, if necessary, improved. The inherent risk of error propagation through various analytical steps in GIS and through the use of models, if unchecked, may lead to completely erroneous results with serious practical implications. The origin of this problem could be traced back to the two stages of information generalization that soil data have to undergo during soil resources inventory: i.e. soil classification and soil mapping. Burrough (1989) pointed out that the quality, and hence the usefulness, of the maps that have been derived by reclassifying and recombining the soil map units with each other or with other mapped data is governed by the quality of the basic data and by the ways in which data are encapsulated and stored (Burrough, 1998; Webster, 1968). Research into the spatial variation of soil and the errors associated with the field and laboratory estimates of soil properties has accumulated much evidence to suggest that the simple model of rigidly defined uniform building blocks embodied in most soil classifications, such as Soil Taxonomy, and also in the areal units mapped by conventional soil survey, produces an incomplete and sometimes unsatisfactory or even misleading description of the landscape.

4 HYPOTHESES

H_{01}: There are no significant differences in accuracy of estimates of land performance, as predicted by suitability class and measured by crop yields, between suitability maps derived from the current soil information paradigm based on generalization (i.e. soil classes and polygons) and a new paradigm based on retaining ungeneralized "hard" point-data, optimal spatial interpolation techniques and the "Fuzzy" representation of soil boundaries.

H_{A1}: There are significant differences in accuracy of estimates of land performance, as predicted by suitability classes, derived from the current paradigm of soil information and a paradigm based on ungeneralized data, interpolation and Fuzzy boundaries.

H_{02}: Fuzzy boundary representation does not significantly improve the accuracy of land performance, as predicted by suitability classes and measured crop yields, over the accuracy obtained from optimal spatial interpolation techniques alone.

H_{A2}: Fuzzy boundary representation significantly improves accuracy of predictions of land performance, as predicted by the suitability class and measured by crop yields, over the accuracy of estimates derived from optimal spatial interpolation alone.

5.0 METHODS AND ANALYSIS

5.1Proposed Paradigm for Providing Soil Information for Land-Use Planning

The methodology advanced by this thesis aims at examining the practical implications of the paradigm shift from the conventional mapping and interpretation of soil/land information, as used in current land evaluation and suitability interpretations, to a paradigm that utilizes a different way of dealing with spatial variability and with uncertainty in mapping. Geostatistical prediction methods can be used in order to give an unbiased spatial prediction. Such a predictor has the capacity to generate continuous multiple coverages of individual soil properties over the study area. Many different techniques are available for spatial prediction. These essentially involve some form of spatial interpolation of point data. The different algorithms of the Kriging technique are being used extensively by soil scientists as a spatial predictor. and hence for reconstruction of the continuum of spatial variability of a given soil property over and area, starting from point-data (Ponce-Hernandez, 1994).

On the other hand, Fuzzy Set Theory applied to soil classifications is useful for creating continuous classes that take into account the transitional nature of soil variability. Odeh et al. (1992) applied a combined approach that uses Fuzzy Set calculus to generate the continuous soil classes, and a

Kriging technique for optimal interpolation of the coefficients that represent the degree of membership to a given class.

There is sufficient evidence in published work to believe that a paradigm, consisting of retaining ungeneralized or "hard" point data in the databases, and a set of algorithms for optimal spatial interpolation and Fuzzy boundary representation of a given soil property, would allow for providing thematic coverages (one property at a time) from the database upon request, doing away with the need for classification and choropleth mapping and all their shortcomings.

5.1.2 Ungeneralized Point Data

Point data refer to sets of values of soil or site properties obtained from the small areas occupied by a pedon, or to the small areas from which soil samples were bulked before analysis.

The scientist cannot record what the soil is like everywhere: he or she can at best measure properties, whether directly in the field or on material taken into the laboratory, of a small portion of the mantle; that is, from a sample. If we want to know what the soil of any area is like, we must be satisfied with measurements made on a part of it, that is, on a sample. Soil also varies from place to place, often very considerably, so measurements of the soil at one sampling site cannot be used to describe all the soil. Usually, it is more meaningful to use averages to describe the soil of each region separately. When a soil survey is carried out for planning land-use, the sampling sites and maps produced from them are usually intended to enable land managers to predict values of soil properties at sites that have not been sampled, and to supply soil information to the land user in order to guide future land-use decision-making.

Conventional soil maps have been used with considerable success in this way where the soil surveyor has had a good understanding of the land-use decision likely to be made on the basis of this map. However, changes in agriculture and the introduction of novel land uses mean that, increasingly, future land use decision makers will make demands on a soil map that the surveyor could not have anticipated. Under these circumstances not only would land use capability maps compiled based on soil survey be irrelevant, but also the soil map itself would often be of seriously limited usefulness. These limitations would arise either because observations required by the eventual user were not made in the first place, or because the observations that were made were not incorporated in the classification scheme used in the soil map.

While there is no real alternative to resurvey in the first instance (where necessary observations have never been made), many problems arise simply because a soil map is an inadequate data retrieval system for the volume of point information collected in a soil survey. Typically even a quick auger observation involves recording of a number of soil properties (color, texture, etc.) over a range of measured depths. The total information content (even if observations irrelevant to a particular area are excluded) is unlikely to be much less than 100 bits per observation point and it frequently may be much greater. Yet, on a conventional soil map this information must be reduced by assignment of a single mapping unit. The number of mapping units on a soil map is usually less than 100 and so the information content for any point on the map is at most 6-7 bits. The soil map thus contains less than 10% (and often much less) of the point information generated by the original survey (Giltrap, 1980). However, such information is of little practical value unless we can use it as a reliable description of the area as a whole. We want information that is truly representative of the area, and means of sampling that will ensure this, bearing in mind that soil is very variable.

Conventionally, in order to model natural phenomena in terms that people can understand, it has been always necessary to abstract and to generalize. Because most natural phenomena are complex, varying at many scales in space and time, scientists have been forced to select the most important aspects of any given phenomenon and to use these as the basis for information storage and transfer. Generalizing and abstracting complex, multiscale phenomena requires serious thought: it is far from easy. The best generalization for one purpose may be unsuitable for another. Ideally, each discipline and each scientist would derive a separate generalization for every different situation as the need arises (Burrough and Frank, 1996).

In order to characterize and portray soil variability, land/soil resources surveys have used two important tools: generalization and classification. Soil classification uses crisp discrete classes and every soil in the area falls into a class. Class limits are sharp and discrete, yet by imposing an arbitrary breadth of class and by letting the class be represented by a central concept (typical profile) the original field information is generalized. The second stage of generalization involves the determination of the spatial extent of the soil class. In reality, since most soil changes are not abrupt, the mapping of the continuous variation of the soil landscape into parcels of land with discrete boundaries involves information generalization and the inclusion of soils from other classes in the form of impurities (Beckett, 1984; Ponce-Hernandez, 1994).

Without a computer it has been often difficult to handle the volume of data produced and to store them properly. An alternative to the discrete polygon data model for soil is to assume that soil properties vary gradually over the landscape. The soil is sampled at a series of locations and attributes are determined for these samples. The alternative procedure would be to focus on retaining ungeneralized hard point-data, since the computer storage capacity and technology for storing and manipulating the large volumes of data generated from field survey now exists. Then, one would create surfaces that represent the continuous spatial variation of soil properties by spatially interpolating these hard point-data. The retention of ungeneralized point-data avoids the two generalization stages in classification and mapping (Beckett, 1984; Ponce-Hernandez, 1994,1995; Burrough, 1996).

5.1.3 Point Data in Geographical Information Systems (GIS)

Until recently, all spatial data were stored and presented to the user in classified form on paper maps. Developments in computer technology now enable land use planners to analyse mapped data and to link them to other relevant information so that all kinds of questions related to the position of an object in relation to other objects can be ascertained. The new GIS technology also makes it possible to evaluate various scenarios (i.e. possible plans) before they are carried out. The tools that provide these new opportunities for creative planning are now part of Geographical Information Systems (GIS). GIS are the result of developments in a number of related sciences, including computer-aided design, computer-assisted cartography, remote sensing, spatial statistics and database technology (Burrough, 1987). Their development has been the result of a marriage of new technology with the basic requirements of a planner who wishes to be able to use all available data to the full. Basically, a modern GIS stores spatial data about soil, land use, climate and so on, in terms of basic graphic entities such as points, lines and areas (polygons). Sets of attributes describing the values of properties that apply to the whole entity are held in an associated relational database.

The spatial distribution of the points, lines or areas may be represented in either the raster (grid cell) or the vector formats. Data at individual point locations (soil profile pits) can be given a spatial extent. in the form of polygons, in several ways. These are:

a) by reclassifying the area entity in which the point falls;

b) by numerical interpolation and threading of contours; and

c) by using a function(e.g. as exponential decay) to model variation over space.

The vector data model represents space as a series of discrete entity-defined point, line or polygon units which are geographically referenced by Cartesian Coordinates. Simple points, lines and polygon entities are essentially static representations of phenomena in terms of XY coordinates. They are supposed to be unchanging, and do not contain any inform ation about temporal or internal spatial variability. A point entity implies that the geographical extents of the object are limited to a location that can be specified by one set of XY coordinates at the level of resolution of the abstraction. The attributes of entities may be expressed by Boolean, nominal, integer, or real data types. In GIS the primitive entities are points, lines, polygons, and pixels (grid elements). Complex entities are defined in terms of their geographical location (spatial coordinates or geometry), their attributes (properties) and relationships (topology). For example, delineation on a soil map can be represented as an area entity (the set of XY coordinates defining the boundary of the enclosing polygon). The associated attributes will be the soil properties as defined in the map legend. and related topological information may indicate the kinds of soil that are adjacent to the delineation.

Besides providing facilities (a) for entering all the necessary spatial and attribute data, (b) for maintaining a digital database and (c) for extracting and displaying information from the database, a GIS can provide a wide range of options for transforming the data according to a user's requirements. The options allow the user to operate on the spatial and attribute data separately. Points are used to represent the location of geographic phenomena at a point or to represent a map feature that is too small to be shown as an area or line. In a vector-based GIS, the identification of the points and lines contained within a polygon area is a specialized search function. In a raster-based GIS, it is essentially an overlay operation, with the polygon in one data layer and the points and/or lines in a second data layer. A neighbourhood operation permits a point to be considered in terms of its surroundings. For example, a value of an attribute at an unsampled point can be estimated from surrounding observations. A moving window can be placed over a set of points and the mean, maximum, minimum, range or index of diversity can be estimated from the data therein. The rate of change of a continuous function at the point can be estimated (e.g. estimates of gradient and aspect of slope from a digital terrain model of the hypsometric curve), as can the steepest downhill path. A buffer zone can be generated around the point to a given distance, spreading out either isotropically over a given surface (such as a landform), or through given barriers, such as those caused by natural features or by other constraints (Burrough 1987,1989, 1998; Aronoff, 1995).

5.1.4 Spatial Variability of Soil and Regionalized Variable Theory

Variation of soil properties from point to point in the landscape is derived from many causes. Climate produces gradual changes over large distances and variations of soil properties induced by climate are different from one zone to another. Climatic regions are especially useful for regional-scale land evaluation where the broad climatic differences are of overriding importance. Maps based on systems modelling regional variations are usually at small scale. For instance, there are soil variations that can be attributed to climate. Soils from temperate regions may be contrastingly different to those from semi-arid regions. Typically, organic matter content is low in soils of arid regions and so is their organic N content (London, 1991). This contrasts with the same parameters in temperate regions.

Parent materials affect soil distribution at two scales, small and large. The former is that of broad types of parent material. Soil may vary irregularly over short distances, as in the main types of drift material. Kantey and Morse (1965) and Robinson and Lloyd (1915) point out that soils formed on transported materials tend to be more variable than those weathered from bedrock.

Relief, with its associated influence on hydrology, produces the detailed distribution patterns which dominate soil maps at medium to large scale soil surveys. Within the soil itself some physical and chemical processes tend to increase lateral variability. Many biological activities increase local variability, for example the localised uptake of nutrient and water, or their concentration beneath the tree canopy (Beckett and Webster, 1971).

In general, the change in spatial variability with increasing scale factor depends on the soil factors determining spatial change (Wilding and Drees, 1983). Total variance will increase as sampling area increases (Beckett and Webster, 1971), but relative contributions of variance at different scales to the total variance follow no consistent pattern (Wilding and Drees, 1983). Much of the variability for some properties may occur over short distances within sampling units (Dent and Young, 1981).

A review of soil variability by Beckett and Webster (1971) revealed that up to 50 percent of the variability between similar soils occurred within 1 m, a clear indication of the differences which can occur within individual pedons. In their study of changes in soil properties over a range of sampling distances, Webster and Butler (1976) found that most of the within field variance of phosphorous occurred within a distance of 5 m; of bulk-density and water content over 18 m; of soluble potassium over 56-180 m; of pH, over 56-180 m; and morphological properties over 50 to 180 m. Variability (Coefficients of Variation) of electric conductivity (EC) and Sodium

Adsorption Ratio (SAR) in 1.44 ha of a cultivated field was 79% and 81%, respectively. In general, soil chemical properties show greater spatial variability than physical and mineralogical properties (Ponce-Hernandez, 1995). Soil variability is of the utmost relevance to soil mapping. Even when mapped in detail, only about one tenmillionth of the soil body is actually examined. A perfect soil map is impossible (Burnham, 1986). When mapping soils in a complex area with a mixture of taxonomic units, it is expected that the taxonomic units with more than 10% of the characteristics of the central concept pedon will be mentioned in the map key and report. These inclusions can be expected to account for 70% to 95% of the area being mapped. Sometimes soil series are used, both as mapping units and as low level taxonomic units. It is important to remember the distinction. Many soil series used as mapping units will be made with up to 80-90% of that series'central concept in the taxonomic sense, with 10-20% of other series as inclusions or impurities (Burnham, 1986). However, in practice, some soil series mapping units may contain no more than 50-60% of the central concept of that series (Ragg and Henderson, 1980). Recognition of the importance of soil spatial variability has led to the study of soil heterogeneity ranging from local scale to the global scale (FAO, 1974).

5.1.5 Regionalized Variable Theory

Retaining point data represents a very successful alternative to conventional mapping to examine the nature of the spatial variation from values at data points before any interpolation is carried out. This has been precisely the approach of Geostatistics which considers all attributes to vary throughout the two-dimensional or three-dimensional space as Regionalized Variables. As far as soil science is concerned, some of the most promising and exciting developments, however, have taken place only since the last decade with the application of Regionalized Variable Theory (Matheron, 1971). This theory enables the spatial dependence in a property to be estimated quantitatively from data, under reasonable assumptions, and then to be used to estimate means with minimum variance (McBratney and Webster, 1983).

Regionalized Variable Theory assumes that the spatial variability of any variable can be expressed as the sum of three major components. These are: (a) a structural component, associated with a constant mean value or a constant trend or drift; (b) a random, spatially correlated component, known as the variation of the regionalized variable; and (c) a random noise or residual error term.

Let χ be a position in 1,2 or 3 dimensions. Then the value of a random variable Z at χ is given by

$$\boldsymbol{Z}\,(\chi) = \boldsymbol{m}\,(\chi) + \acute{\varepsilon}\,(\chi) + \breve{\varepsilon}, \qquad (1.0)$$

where $\boldsymbol{m}\,(\chi)$ is a deterministic function describing the structural component of $\boldsymbol{Z}$ at χ; $\acute{\varepsilon}\,(\chi)$ is the term denoting the stochastic, locally varying but spatially dependent residuals from $\boldsymbol{m}\,(\chi)$- the regionalized variable; and $\breve{\varepsilon}$ is a residual, spatially independent Gaussian noise term having zero mean and variance σ^2.

In order to investigate each of these components of spatial variability, the first step is to decide on a suitable function for $\boldsymbol{m}\,(\chi)$. In the simplest case, where no trend or drift is present, $\boldsymbol{m}\,(\chi)$ equals the mean value in the sampling area, and the average or expected difference between any two places χ and $\chi+\boldsymbol{h}$ separated by a distance by the vector $\boldsymbol{h}$, will be zero:

$$E\,[z\,(\chi) - z\,(\chi + h)] = 0, \qquad (1.2)$$

where $z\,(\chi)$, $z\,(\chi + h)$ are the values of random variable z at locations χ and $\chi + \boldsymbol{h}$. Also, it is assumed that the variance of differences depends only on the distance between sites, $\boldsymbol{h}$, so that the mathematical expectation is:

$$E[\{z(\chi) - z(\chi + h)\}^2] = E[\{\acute{\varepsilon}(\chi) - \acute{\varepsilon}(\chi + h)\}^2] = 2\gamma(h), \qquad (1.3)$$

where $\gamma(\boldsymbol{h})$ is known as the semivariance. The two conditions (i.e. stationarity of difference and variance of differences) define the requirements for the so-called intrinsic Hypothesis of Regionalized Variable theory. This means that once structural effects have been accounted for, the remaining variation is homogeneous, so that differences between sites are merely a function of the spatial covariance structure and the distances between them. If conditions specified by the intrinsic hypothesis are fulfilled, the semivariance can be estimated from the sample data:

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}\{z(x_i) - z(x_i + h)\}^2 \tag{1.4}$$

where $\boldsymbol{n}$ is the number of pairs of sample points of observations of the values of attribute z at position x separated by distance $\boldsymbol{h}$.

A plot of $\gamma(\boldsymbol{h})$ against $\boldsymbol{h}$ is know as the *experimental semi-variogram* **(Fig.2)**. The experimental semi-variogram is the key to providing a quantitative description of the spatial covariance structure of the attribute, which provides useful information for interpolation, optimizing sampling and determining spatial pattern. The semi-variogram is quite useful as a tool for elucidating the nature and structure of spatial variability in a given regionalized variable. A model can be fit to such variation allowing for quantitative representation of spatial variability. This information can then be used for estimation of values across the space by incorporating it in the spatial interpolation method. The semi-variogram is central to geostatistics and the single most important tool in geostatistical applications to soil (McBrateny and Webster, 1986). Its accurate estimation is critical in the success of spatial interpolation and the generation of raster maps (Ponce-Hernandez, 1994). Semi-variogram analysis has the added advantage of defining parameters for local estimation by Kriging. **Fig.2** shows a typical experimental variogram of data from a varying attribute, such as a soil property. The curve that has been fitted through the experimentally derived data points displays several important features. First, at large values of the lag, ***h***, it levels off. This horizontal part is known as the ***sill (C)***; it implies that at these values of the lag there is no spatial dependence between the data points because all estimates of variances of differences will be invariant with sample separation distance. Second, the curve rises from low value of $\gamma(\boldsymbol{h})$ to the sill, reaching it at value of ***h*** known as the ***range***. This is the critically important part of the variogram because it describes how inter-site differences are spatially dependent. Clearly, if the distance separating an unvisited site from a data point is greater than the range then that data point can make no useful contribution to the interpolation; it is too far away (Burrough, 1998). Semi-variogram ranges depend on the scale of observation (i.e. the minimum lag distance) and the spatial interaction of soil processes affecting each property at the sampling scale used (Trangmar et al., 1985). Semi-variances may also increase continuously without showing a definite range and sill, thus preventing definition of a spatial variance and showing the presence of trend effects and nonstationarity (Webster and Burgess, 1980). Ideally, the experimental semi-variogram should pass through the origin when the distance of sample separation is zero. However, many soil properties have nonzero variance for semi-variances as h tends to zero **(Fig 2)**. This nonzero variance is called the nugget effect (C_0) (Journel and Huijbregts, 1978; Isaaks and Strivastasa, 1989). The experimental semi-variogram exhibits pure nugget effect (100% of sill) when $\gamma(h)$ equals the sill at all values of h. Pure nugget effect arises from very large point-to-point variation at short distances and indicates a total absence of spatial correlation at the sampling scale used. Increasing the detail of sampling will often reveal structure in the apparently random effects of the nugget and pure nugget variances (Burrough, 1983). It is usual to fit a model to the discrete sample semi-variances because the true variogram of a region is continuous and the estimates are subject to error, especially if the sample is small, which may make the variogram appear erratic (Oliver and Webster, 1990).

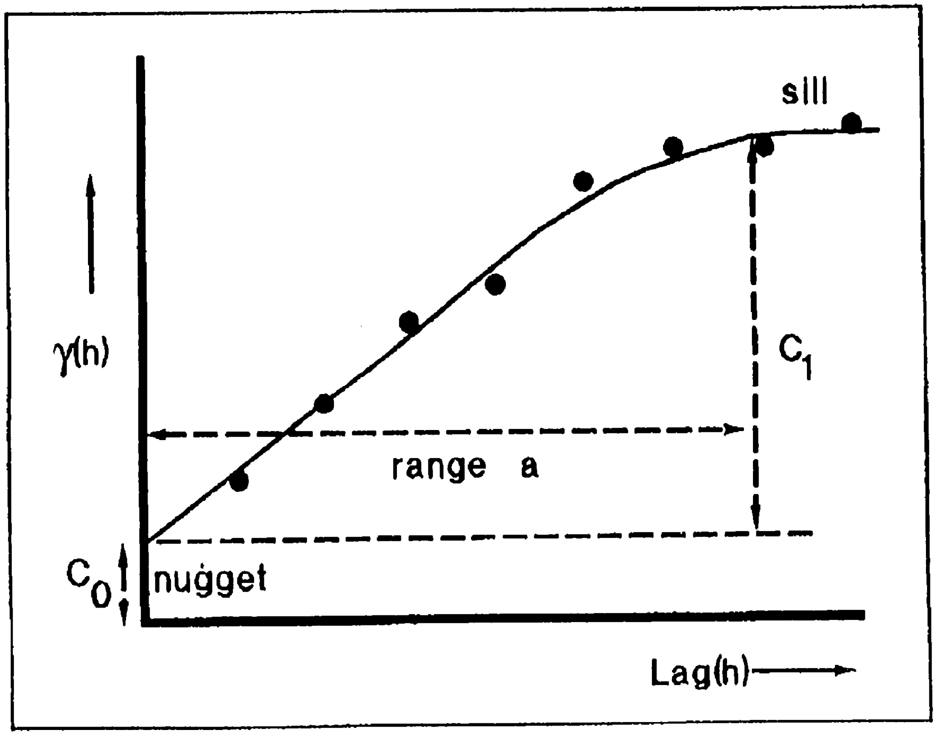


Fig 2. An example of simple experimental transitional variogram with range, nugget, and sill (Source: Burrough and McDonnell, 1998)

Spherical, exponential, linear and Gaussian models **(Fig.3)** are commonly used as best-fit functions through the scatter of points in the semi-variogram. A variogram that can be fitted by a Gaussian variogram model indicates a smoothly varying pattern, such as often occurs with elevation data, and it is often used to model extremely continuous phenomena. A variogram modelled by a spherical variogram, which is the most commonly used, has a clear transition point which implies one pattern is dominant. The choice of an exponential variogram may suggest that the pattern of variation shows a gradual transition over a spread of ranges or that several patterns interfere with one another. The exponential model is linear over very short distance near the origin; however, it rises more steeply and then flattens out more gradually. The linear model is not a transition model since it does not reach a sill, but increases linearly with h in a non-bounded way. It is important to choose the appropriate model for estimating the semi-variogram because each model yields quite different values for the nugget variance and range, both of which are critical parameters for Kriging (Burrough, 1998; Issaks and Srivastava. 1989) All above models describe isotropic variation. But soil does not vary equally in all directions. There are numerous situations where the variation is anisotropic. In this situation each direction has its own semi-variogram differing from those in other directions. For example, soil properties are isotropic if they vary in a similar manner in all directions and one semi-variogram applies to all parts of the study area. Geometrical anisotropy occurs when variation for a given distance ***h*** in one direction is equivalent to variation over a distance ***kh*** in another direction. The anisotropy ratio ***k*** indicates the relative size in which directional variation is elongated in the direction of minimum variation. The direction of maximum variation is assumed to occur perpendicular to the direction of minimum variation (David, 1977). The anisotropy ratio would equal 1 and define a circular zone of influence if variation was the same in all directions, i.e. isotropic (Trangmar et al., 1985). If the sample pattern is noticeably anisotropic, with the sample spacing being much smaller in some directions than on the others, the distance parameters will depend on the direction in which the anisotropy is present. The utility of anisotropic modelling lies in identification of changes in spatial dependence with direction, which in turn reflects soil-forming processes (Trangmar, 1985). However, soil properties which are highly correlated and whose semi-variograms vary anisotropically often have anisotropic cross-semi-variograms, i.e. a semi-variogram that accounts for the strength of association between the two properties (McBratney and Webster, 1983).

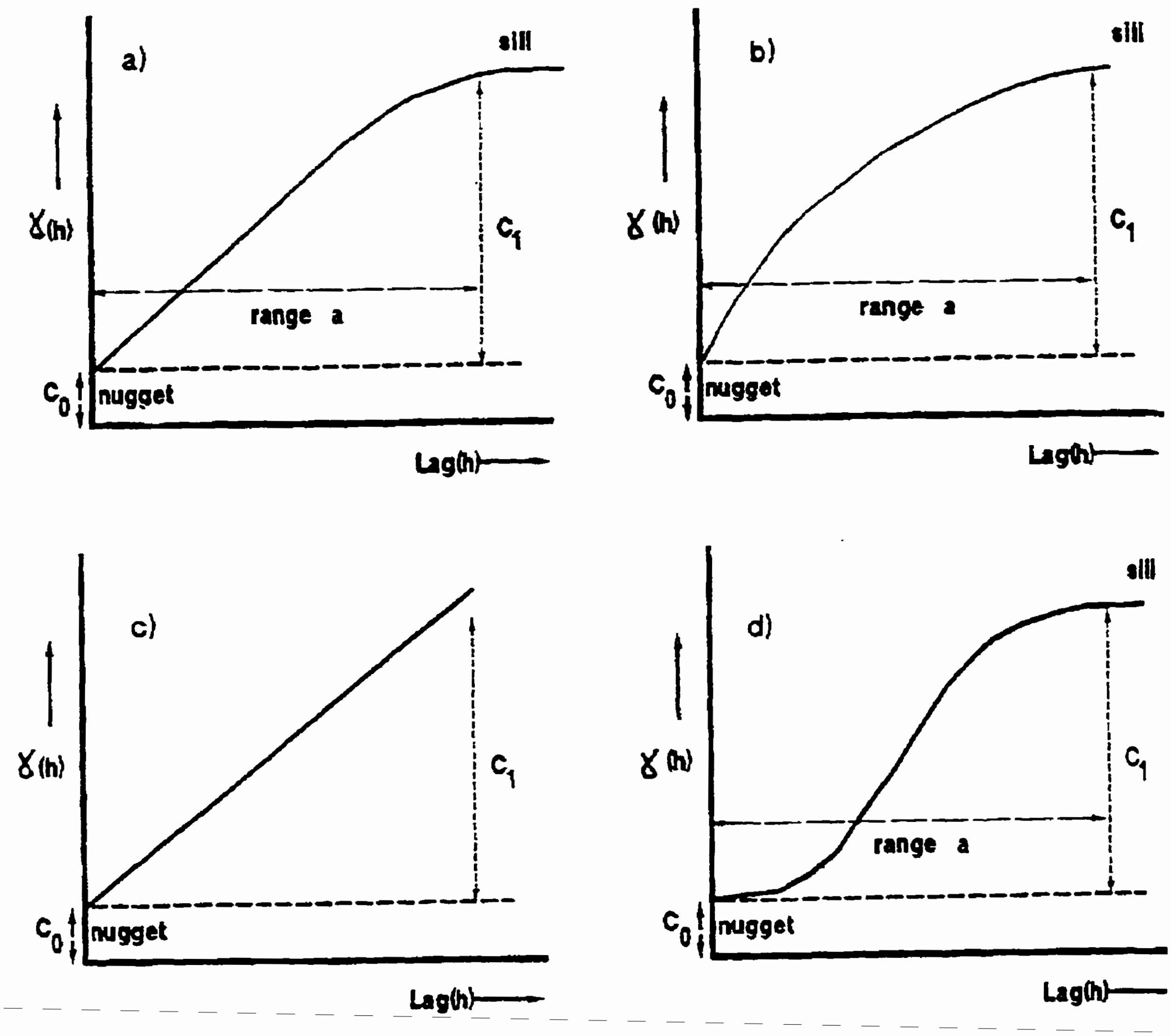


Fig 3. Examples of the most commonly used variogram models: **(a)** spherical; **(b)** exponential; **(c)** linear; and **(d)** *Gaussian* (Source : Burrough and McDonell, 1998)

5.1.6 Spatial Interpolation (estimation)

Interpolation is the procedure of predicting the value of attributes at unsampled sites from measurements made at point locations within the same area or region. Interpolation is used to convert data from point observations to continuous fields so that the spatial patterns sampled by these measurements can be compared with the spatial patterns of other spatial entities. So spatial interpolation, in a sense, is a process of spatial estimation. Interpolation is necessary:

(a) When a discretized surface (i.e. a surface resulting from the division of an area into a set of regular or irregular tiles or "thessellations") has a different level of resolution than what is required. In this case interpolation is needed to create the tiles or pixels at the new resolution required. This function is often called "resampling" in GIS software. An example of this is the conversion of scanned images (aerial photographs or remotely sensed images) from gridded tessellations, with a given size and/or orientation, to another, or

(b) When a continuous surface is represented by a data model that is different from that required, or

(c) When the data we have do not cover the domain of interest completely (i.e. they are samples) (Burrough, 1998).

The methods of interpolation can be divided into two groups, called global (e.g. trend surface models) and local interpolators (e.g. Thiessen polygons, inverse distance, bi-cubic splines, Kriging) depending on the structure of the interpolating function. Global interpolators use all available data to provide prediction for the whole area of interest, while local interpolators operate within a small zone around the point being interpolated to ensure that estimates are made only with data from locations in the immediate neighbourhood and fitting is as good as possible. The chief global approach is trend surface analysis, whereby polynominal or sometimes trigonometric functions are fitted by least squares regression on the spatial coordinates as predictors. This simple approach has several shortcomings: it loses detail because of powerful smoothing; instability may be caused by outliers or observational errors or when enough terms are included in the function to retain local detail; and variation in one part of the region affects the fit of the surface everywhere (Oliver, 1990). Therefore, the global approach is only useful for describing broad geographical trends (cf. Burrough et al., 1977).

When data are abundant, most interpolation techniques give similar results. When data are sparse, however, the assumptions made about the underlying variation that has been sampled and the choice of method and its parameters can be critical if one is to avoid misleading results (Burrough, 1998).

Since the soil scientist has been concerned with spatial variation of soil, geostatistical methods of interpolation, popularly known as Kriging in various forms, are being used extensively as a spatial predictor. Kriging attempts to optimize interpolation by dividing spatial variation into three components: deterministic variation, spatially autocorrelated variation, and uncorrelated noise. The character of spatially correlated variation is encapsulated in the descriptor function such as the semi-variogram. Since our main objective in recognizing the variability over the study area is to estimate soil properties for land-use planning purposes, Kriging (block Kriging) appears to have all the attributes desired for modelling the continuum of soil variability and estimating values for the study area. This can be done by interpolating at grid cells of a specific size. For quantitative spatial modelling in this thesis, Kriging will be used. The technique clearly fulfils the requirements in our research plan. The development of Kriging has been due largely to the development of the main body of Regionalized Variable Theory (section 4.3.1). The term Kriging embraces a set of methods for local estimation, including simple and ordinary Kriging, co-Kriging, universal Kriging and disjunctive Kriging. Simple point estimation is probably the most common Kriging procedure used in soil science (Trangmar et al., 1985). Given that the spatially dependent random variations are not swamped by uncorrelated noise, the fitted variogram can be used to determine the weights λ_i needed for local interpolation. The interpolated value $(\hat{Z})$ of a regionalized variable Z at location χ'_0 can be calculated from:

$$\hat{Z}(\chi_0) = \sum_{i=1}^{n} \lambda_i Z(\chi_i) \quad (1.5)$$

where n is the number of neighboring samples $Z(x_i)$ and λ_i are weights applied to each $Z(x_i)$ The weights are chosen so that the estimate $Z(x_0)$ of the true value $Z(x_0)$ is unbiased, i.e.:

$$E[\hat{Z}(\chi_0) - (Z(\chi_0)] = 0 \quad (1.6)$$

And the estimation variance σ_k^2 is minimized, i.e.,

$$\sigma_k^2 = \mathrm{VAR}\ [\hat{Z}(\chi_0) - Z(\chi_0)] = \mathbf{minimum} \quad (1.7)$$

The weights placed on each neighboring sample sum to 1, and their unique combination for which σ_k^2 is minimized can be obtained when:

$$\sum_{i=1}^{n} \lambda_j \gamma(\chi_i, \chi_j) + \mu = \gamma(\chi_i, \chi_0)\ \textit{forall}\ \mathrm{i} \quad (1.8)$$

The values $\gamma(x_i, x_j)$ and $\gamma(\chi_i, \chi_0)$ are the semi-variances, or preferably the covariances (second- order stationarity), between observed locations χ_i and χ_j and between the observed location χ_i and the interpolated location χ_0, respectively. These values are obtained from the semi-variogram of Z. The quantity μ is the Lagrangian multiplier associated with minimization of σ_k^2. Solution of the $n+1$ equations of the Kriging system for each λ_i and μ enables the kriged estimate $Z(x_0)$ to be determined by Eq. (1.5) and the estimation variance to be determined by solving for

$$\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \gamma(\chi_i, \chi_0) + \mu \tag{1.9}$$

In block Kriging, a value for an area or block with its centre at x_0 is estimated rather than values at points. As in punctual Kriging, the Kriged value of property Z for any block V is a weighted average of the observed values χ_i in the neighborhood of the block :

$$\hat{Z}(V) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1.10}$$

The only difference in Eq. (1.10) from the estimation equation (1.5) for punctual (point) Kriging is in the determination of the weighting coefficients. In the weighting procedure, the semi-variances between data points and the interpolated points of punctual Kriging are replaced by the average semi-variances between the data points and all points in the block $[\gamma(x_i, V)]$. The estimation variance is minimized. The minimum estimation variance for block V is

$$\sigma_v^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i, V) + \mu_v - \gamma(V, V) \tag{1.11}$$

where $\gamma(x_i, V)$ is the average semi-variance between the sample points x_i in the neigborhood and those in the block V. $\gamma(V,V)$ is the average semi-variance between all points within V (i.e., the within-block variance of classical statistics), and μ_v is the lagrangian parameter associated with the minimization.

There is often little difference between the estimates from point Kriging and block Kriging, but the block estimates may seem more reliable because as the size of the block increases the estimation variances decrease. Block estimates may also be more realistic since the information from a point is usually intended to represent an area surrounding it. The size of this area or block will depend on the purpose of the survey (Oliver, 1990). The most common use of block Kriging has been for production of isarithmic maps of soil properties. Experience indicates that block Kriging produces smoother maps than point Kriging by interpolating average values over blocks, with the effect of smoothing local discontinuities (Trangmar, 1985). Block Kriging has also been applied to interpolate spatial effects of crop response to variability imposed by soil management practices. Tabor et al. (1984) found that maps of block-Kriged values for nitrate content of cotton petioles indicated a strong response to direction of planting rows and application of irrigation water. Block Kriging of a very fine mesh of grid cells forms the basis of interpolation procedures developed by Giltrap (1983), which provide for prior stratification of the landscape into a number of land classes and can either restrict interpolation to cells within the same land classes or allow interpolation across land classes using a separately calculated autocorrelation function. Using these procedures, Giltrap (1981) was able to produce maps rapidly and cheaply for many soil properties at any scale similar to that of the original interpolation grid.

5.1.7 Geographic Objects with Indeterminate Boundaries: Fuzziness of Soil Boundaries

In the past, scientists and administrators have ignored or suppressed important aspects of inexact or Fuzzy phenomena and for their own reasons have forced them to be objects with crisply defined boundaries, even if such phenomena are more often thought of as continuous fields. Practical scientists, faced with the problem of dividing undivided complex continua have often been forced to discretize. Such is the case in the aerial photograph interpretation of soil or vegetation patterns. The result has been that different scientists have mapped the same area differently (Burrough, 1998). Furthermore, conventional or crisp sets allow only binary membership functions (i.e. True or False): an individual is a member or it is not a member of any given set. The problems associated with binary logic for retrieval can be seen from the following simple example taken from the discipline of land evaluation (Burrough et al., 1992). With binary logic, complex land qualities and suitability classes can be defined using any of the operators AND, OR, NOT, XOR to specify just which combination of attribute values is required for membership in a class for any given purpose. Class membership is commonly defined by specifying the ranges of a certain number of key property values that an individual must meet. To qualify as a member of a given suitability class, an individual point, line or area must match all the specifications of that class. These specifications can be expressed as a multiple Boolean "AND", or intersection, as in a typical GIS overlay operation:.

$$\mathbf{R} = \text{"}\boldsymbol{true}\text{" if } (\,A \text{ AND } B \text{ AND } C \text{ AND } D.....). \qquad (1.12)$$

where **R** is the result and *A*, ***B***, *C*, ***D***,..... represent the specified ranges of the properties. The result is binary, ***true*** being represented by the character **1**, ***false*** being represented by the character **0**. This logic is often extended to a limited number of discrete classes (>**2**) that describe grades of suitability through the principle of most limiting factor (FAO, 1976). For example, if the suitability of a site for a given land use is determined from the levels of several land qualities, then the land quality with the lowest suitability rating (maximum limitation) determines the site classification:

$$\mathbf{R = MIN} \; (\mathit{Q1, Q2, Q3, \ldots\ldots Qn}) \qquad (1.13)$$

where the Qi are the classified values of each key land quality (usually integers).

Consider for example, the problem of measuring the attributes in the field to assess the erosion hazard in gently sloping sites:

IF SLOPE ≥ 10% AND SOIL TEXTURE = SAND AND VEGETATION COVER ≤ 25% THEN EROSION HAZARD IS SEVERE.

This rule specifies a central concept, namely, that bare sandy soils on more than gently sloping sites are prone to extreme erosion. If the rule was used in a GIS on soil polygons with simple attribute values, however, then the data retrieval operation would only find those polygons with an exact match. Clearly, polygons that for one reason or another had attributes that were just outside the class boundaries would be rejected. The result might lead to a serious underestimate of the area of land that is prone to severe erosion (cf. Burrough, 1989). In addition, in existing systems of soil classification, any one individual soil belongs to exactly one class at any one level. Each individual is allocated to a single class, although its proper allocation may be uncertain because of errors in data or vagueness of class definition. No matter how small the differences in properties may be, the allocation of individuals changes abruptly when crossing a class boundary. In this sense the existing classification systems are discontinuous (Mc Bratney et al., 1992).

Soil classifications are often little better than attempts to subdivide a complex continuum into smaller units (Chang and Burrough, 1987). The best approach to a particular soil classification problem will depend to a large extent on the distribution of the soil individuals in some attribute or character space (Butler, 1980). Instead of first classifying observations into exactly defined classes and then averaging the class scores, another strategy would be to rescale the original data on a continuous scale by assigning continuous class membership values. These continuous values could be assigned to individual attributes or to groups of attributes. Individuals that exactly matched strictly defined classes would receive a membership value depending on their degree of closeness to the strictly defined class. Therefore an individual with an observed value just outside the class limits might receive a membership value of 0.95 to indicate that it was not a full member of the class, but should by no means be completely rejected. Individuals that were so far away from the strictly defined classes limits that they receive a membership value of less than 0.5 could be safely rejected.

The problem of dealing with undefined classes and vague boundaries is not unique to soil science but is a part of human experience. Until recently there was no reasonable and quantitative way to handle such imprecision, but Zadeh (1965) introduced his Fuzzy set theory as a means of dealing with inexact concepts. In conventional set theory points are allowed to belong only to one set, whereas in Fuzzy-set theory they may belong totally, partly or not at all to a set. A Fuzzy set is a class of objects with a continuum of grades of membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. Notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of Fuzzy sets are established. In particular, a separation theorem for convex Fuzzy sets is proved without requiring that the Fuzzy sets be disjoint (Zadeh, 1965). A Fuzzy set of attribute values is defined mathematically as follows:

If $X = \{ x \}$ denotes a universe of attribute values (i.e. range of values), then the Fuzzy set A *in* X is the set of ordered pairs

$$A = \{ x, \mu_A(x) \} \qquad x \in X \qquad (1.14)$$

where $\mu_A(x)$ is known as the 'grade of membership of x in A' and $x \in X$ means that x is a value contained in X. Usually $\mu_A(x)$ is a number in the range **0,1** with **1** representing full membership of the set (e.g. the 'representative profile') and **0** non-membership. The grades of membership of x in A reflect a kind of ordering that is not based on probability but on admitted possibility. Put another way, the value of $\mu_A(x)$ of attribute value x in A can be interpreted as the degree of compatibility of the predicate associated with set A and attribute value x. Therefore the value of $\mu_A(x)$ gives us a way of giving a graded answer to the question: "to what degree is a soil profile with attribute value x a member of soil class A?".

The simplest model is given by the following equation which is a general symmetrical bell-form membership function:

$$MF_x = \frac{1}{[1+\{(x-b)/d\}]} \; for\ 0 \leq x \leq p. \qquad (1.15)$$

The parameter ***b*** defines the value of the attribute ***x*** at the central concept or the standard index of the set. The form of the membership function and the position of the crossover points can be easily changed by changing the value of the dispersion index, ***d***. The parameter ***d*** gives the width of the bell curve at the crossover points which defines the transition zone around the central core of the set in the same units as the central concept (Burrough et al., 1992).

Membership functions can be drawn for different soil properties within soil profiles, and such functions can be used in Fuzzy operations to answer simple or complex queries. Land suitability assessments based on a Fuzzy operator with cross-over values representing the critical values of land- use requirements give better results than the strict Boolean approach (crisp boundaries) with operator “OR”, “AND”, etc. (Burrough, 1989). Applications of Fuzzy sets in soil modeling have been explored by, among others, Chang and Burrough (1987), Burrough et al., (1992), Tang et al. (1991), Triantafilis and McBratney (1993), and Davidson et al. (1994). In soil distribution modelling the use of Fuzzy sets and Fuzzy logic was initiated by De Gruijter and McBratney (1988) and McBratney and De Gruijter (1992). For instance, Tang et al. (1991) found that Fuzzy set methods differ from the parametric methods in the use of an explicit weight for the impact of each land characteristic, and in the way of combining the evaluation of each land characteristic into a final suitability class or suitability index. Besides a dominant suitability class, the Fuzzy set method equally provides information on the degree to which the land unit belongs to each of the suitability classes discerned. Davidson et al. (1994) concluded that their study, by applying Fuzzy set methodology in a land evaluation project in Viotia in Greece, yielded more a satisfactory result than the traditional one using the Boolean approach. Finally, Burrough et al. (1992) stated (p.207) "the Fuzzy approach is clearly more flexible than Boolean methods for analysis of land suitability. Because Boolean intersection only accepts sites that match all the strict requirements, many sites are rejected". The only critical issues in the use of Fuzzy set methodology are the choice of membership function, class centers, cross over values and weight. Research on how these can be determined from the data themselves may yield more objective results (Burrough, 1989; Davidson, 1994).

5.1.8 Elements of the Proposed Paradigm

In this research three alternative procedures to the current paradigm for providing and processing soil information are proposed and investigated. These include:

1.Retention of hard-point-data: Storage of ungeneralized point data in the database as they were collected from field survey and lab analysis.

2.Spatial interpolation: Use of geostatistical tools in order to obtain spatial prediction estimates and thematic mapping of individual soil properties on request from reappraising point data. The combination of geostatistical estimation and GIS can provide the interpolating algorithm and tools needed for a new paradigm in acquiring, storing and providing soil/land resources information.

3.Fuzzy Set theory: Application of continuous classification based on constructing a grade of membership on the interpolated classes in order to create smoother and transitional boundaries between the classes.

5.1.9 Methodological Procedures

The methodological procedures used in this thesis were designed for carrying out a comparison of paradigms for representing and providing soil information, and for examining the practical implications of a paradigm shift. The shift in paradigm would be from the conventional discrete classification and mapping of soil/land, as used in current land evaluation, to a proposed new paradigm that utilizes a different way of dealing with the transitional nature of spatial variability of soils. The overall sequence of procedures used in the methods for this research design can be seen in the flow chart displayed in **Fig. 4** and **Fig. 5.**

CURRENT PARADIGM METHODOLOGY

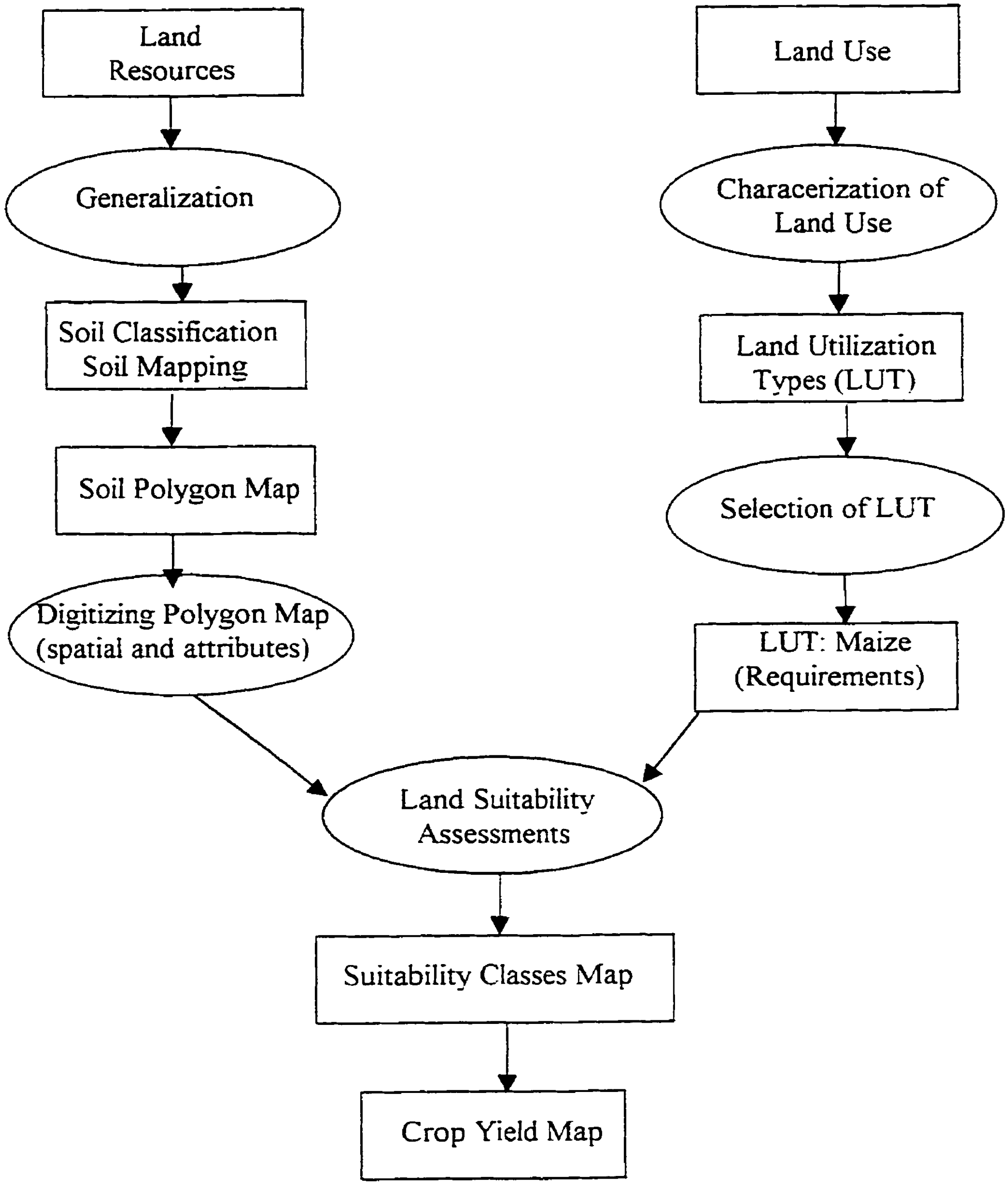


Fig. 4 Current paradigm methodology

PROPOSED PARADIGM METHODOLOGY

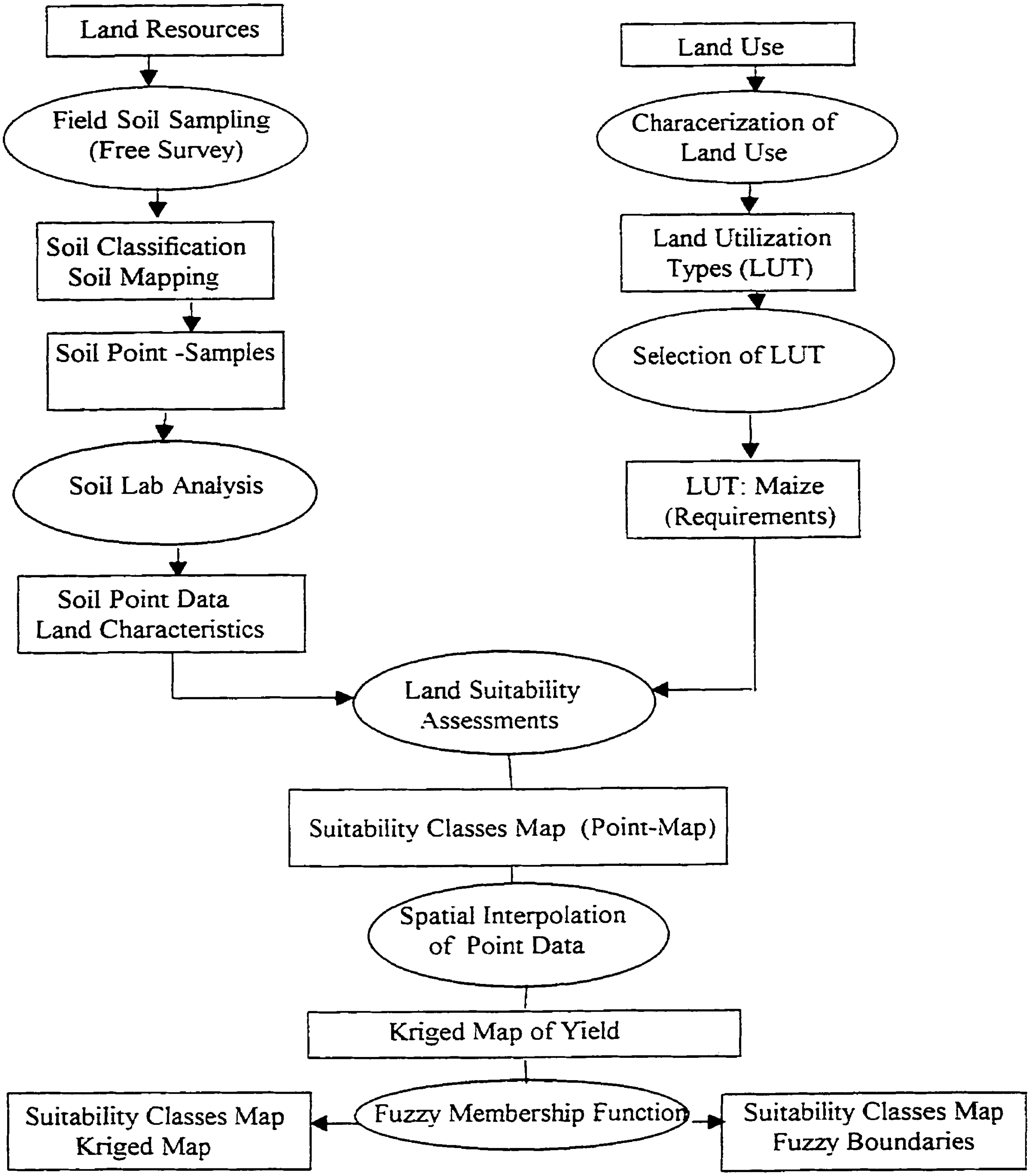


Fig.5 Proposed paradigm methodology

5.2 Description of the Study Area

This study was conducted in the area covered by the Texcoco river watershed in central Mexico. This watershed is located 48 km northeast of Mexico City **(Map 1)**. The watershed is located in Universal Traverse Mercator (UTM) zone 14 and has boundaries within the coordinates of 514000, 2146000 and 530000, 2156000. The watershed is close to 300 km^2 in size and ranges in elevation from 2300 to 4000 m above sea level. The area was selected because data are available within this area.

The climate of the area is characterized by cold and dry winters, while summers are warm with abundant rain. Annual average precipitation varies from 450mm to more than 1500mm depending in physiographic position. More than 80% of all precipitation falls between June and October. Frosts are typically severe from November to February, but generally they begin in October and last well into February. The rainfall pattern is monsoonal. Scattered showers occur from November to May. Substantial showers begin in May and become consistent from June until mid-September (Sandres et al., 1979). The mean annual temperature varies between 12° C and 18° C. The temperature of the coldest month varies between -3 and 18° C and for the warmest month between 6.5 and 22° C (Vargas, 1993).

The watershed has several different zones with varying soil types. Soils are generally medium textured and sandy loams and thus well-suited to maize-based agriculture with hand tools. Soil depth is quite variable. Without taking into account all the differences in soil depth, the soils of Texcoco have some important characteristics for management, such as:

- Soils with incipient structure.
- Friable soil aggregates which are pulverized easily. This makes them susceptible to erosion

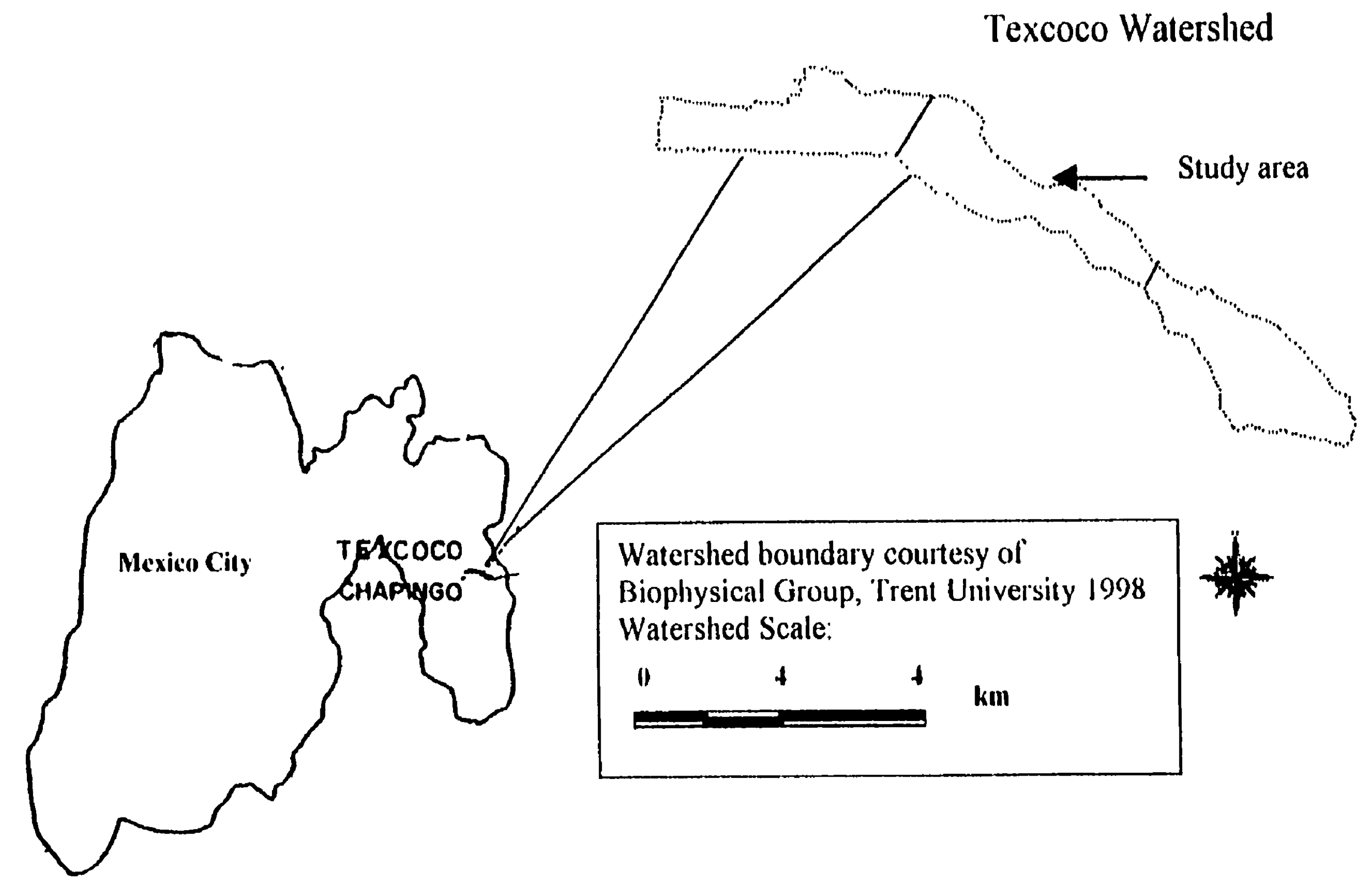
when they are bare.

- Soil with low bulk density values and high total porosity values.

According to INEGI (1986) the main soil units for Texcoco are Zolonchak, Vertisol, Phaeozem, Litosol and Cambisol (Vargas, 1993).

The main crops grown on the irrigated flat lower portion of the watershed are corn, beans, barley, peas, onion, pear, walnut and apple trees. Half of the watershed is covered with crops while the other half is covered with deciduous forest. Corn alone or in association with beans and squash are the most frequently-found crop associations in terms of area covered.

Map 1: Texcoco River Watershed

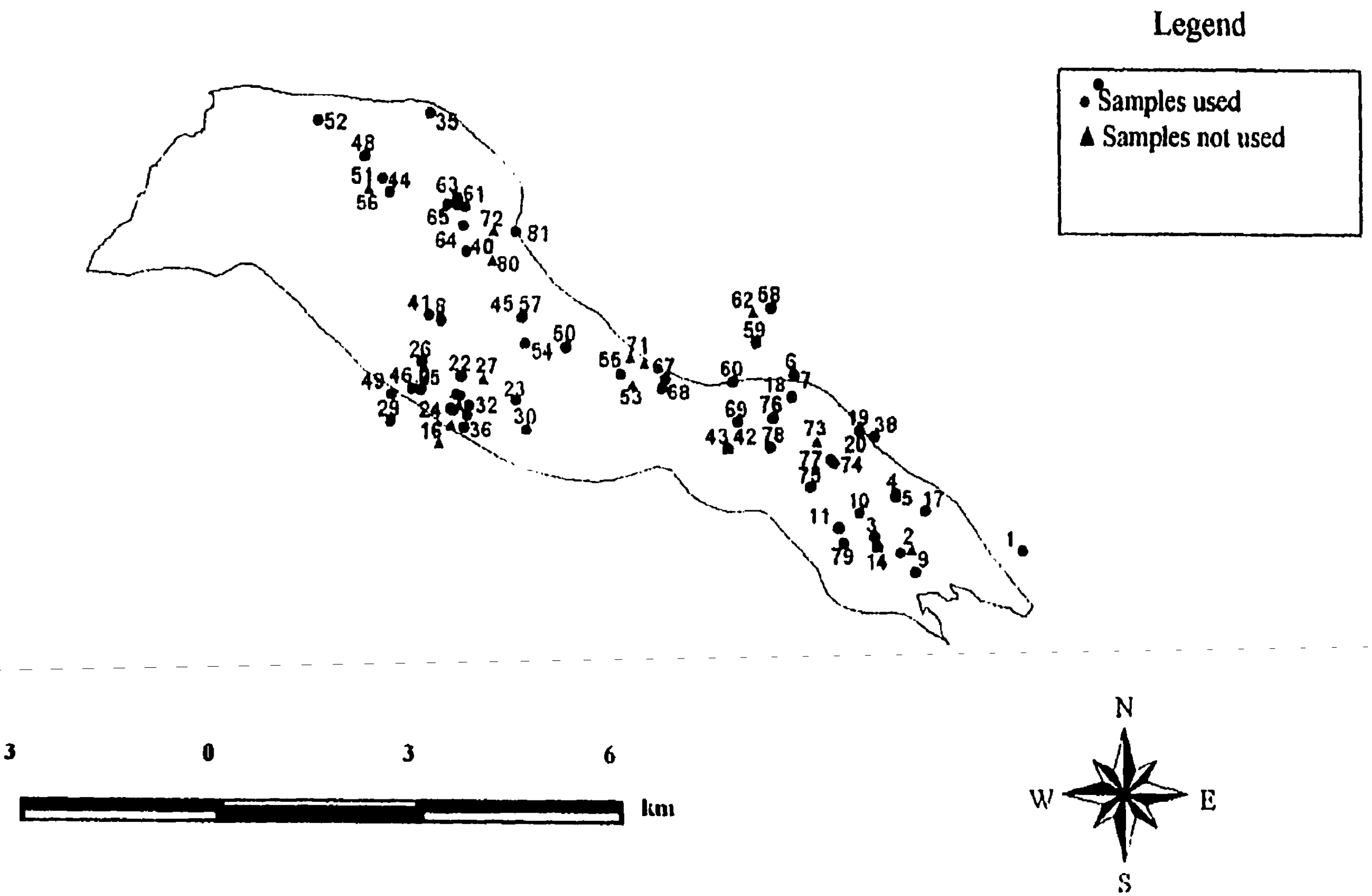


5.3 Sampling Strategy

A soil sampling campaign was undertaken in most of the agriculturally-based area in the watershed **(Map 2)**. The aim of the survey was to sample the spatial variability of the main fertility-related characteristics of the soil in the area. The sampled fields were chosen to be representative of the variability of corn growing conditions in the area. Modifications were made to the initial "free" survey strategy, whenever permission by the farmer was not obtained, or when cooperation with the project not granted. In that sense, the sampling strategy became some form of hybrid between "free" survey and random survey, since other sites were chosen at random to replace the unavailable ones in the general sampling design.

At each point plot, five soil samples were taken and mixed to produce a composite bulk sample representative of the plot. Soil samples were air dried, passed through a 2 mm sieve and stored in plastic bags, ready for laboratory analysis (Trent University). The sampling design yielded a total of 83 samples. It is worth mention that only the central portion of the watershed was sampled. This is due to the fact that this is the portion of the watershed that is intensively cultivated under rainfed agriculture, and this type of agriculture is one of the main concerns of this thesis. Therefore this study concentrates on the middle portion of the watershed where the most important land use is rainfed agriculture, and the land has slopes that vary from gentle to flat.

Map 2 Point –samples (83) for the Study area
(part of the Texcoco River Watershed)



5.4 Sample Treatment and Laboratory Analysis

At each sampling site the following measurements were made to create a data- base of site characteristics: slope, altitude, depth and coarse fragments **(Appendix 1a)**. The slope was measured with a clinometer, depth to the hardpan was measured with a soil auger and tape, coarse fragments were estimated as a percentage value, and the altitude was measured using a calibrated altimeter. A **GPS** (Global Position System) was used to record the coordinates of each sampling site and to check the altitude value given by the altimeter. The samples were taken to the laboratory and the soil was analyzed for: **pH, organic matter content (OM), exchangeable cations** such as **potassium (K), magnesium (Mg), calcium (Ca), Cation Exchange Capacity(CEC), electrical conductivity (EC)** and **soil texture** or particle size distribution **(Appendix 1b and c)**.

The **pH** of each sample was determined in water. A 5 g sub-sample of sieved soil was added to 20 mL of distilled deionized water in a paper cup, mixed several times over a 15 minute period with a plastic stirrer and left to settle for 45 minutes. The pH of the clear liquid was then measured using a Corning 135 pH/ion meter with a glass electrode.

The **Organic matter** of the soils was estimated by the loss of weight on ignition (LOI), according to Karla and Maynard (1991). A 1 g sub-sample of sieved soil was weighed in a crucible and then ignited at 400 ^{0}C for 12 hours in a muffle furnace before cooling and reweighing. The weight of the sample after ignition allowed for the calculation of LOI as follows:

$$\textbf{Loss on ignition(\%OM)} = \frac{\textbf{weight of sample before ignition- weight after ignition}}{\textbf{Weight of sample before ignition}} * 100$$

The procedure was completed three times for each sample. and the mean of three measurements per site was taken as **% OM**.

Exchangeable cations and cation exchange capacity were determined by manual leaching using vacuum extraction according to Karla and Maynard (1991). The method involves leaching soil with a buffered (pH 7.0) 1.0M ammonium acetate solution in which the displaced exchangeable cations were measured. The exchange capacity was filled with ammonium as the soil and solution was left to stand overnight. Then the leachate was filtered from the soil solution using gentle suction. An aliquot was saved for determination of **K, Mg** and **Ca** concentrations using atomic absorption spectrophotometry. The excess ammonium acetate was then displaced by ethyl alcohol. Acidified sodium chloride then displaced exchangeable ammonium which was measured to yield a quantity for the **CEC**.

Electrical conductivity was determined by saturation extract (Black, 1965). Distilled water was added to 200 g of soil until a paste was formed (there was a shiny surface but no standing water). After allowing the saturated soil paste to stand 4 hours, the paste was filtered by using a large filter apparatus and vaccum pump. 1 drop of 0.1% sodium hexametaphosphate solution was added to the filtrate. A calibrated conductivity meter (probe) was then inserted into the filtrate to record the electrical conductivity.

Soil texture was determined by hydrometer analysis (McKeague, 1978). In this method the mineral part of the soil was separated into different-sized fractions (sand at >50 μm, silt between 50 and 2 μm and clay < 2 μm).

The **% of Base Saturation** was calculated from the following equation as the proportion of the **CEC** accounted for by exchangable bases (**Ca, Mg, K** and **Na**) (London,1991):

Percentage base saturation = 100* exchangable bases/CEC

A widely used measure of Na levels in soil is the exchangable Na percentage **(ESP),** which is defined as (London, 1991):

ESP = (Exchangable Na/ CEC) * 100

5.5 Spatial Database Development

The Integrated Land and Water Information System **(ILWIS)** was used to digitize three cover maps of: the watershed boundary map **(Map1)**, study area showing sample sites **(Map 2)** and soil polygons **(Map 3)**. The boundary of the watershed was delineated using a 1:20000 ortho-photomap. The boundary delineation was based on a previous delineation in the study area with some modifications, using aerial photographs. A base map at a scale of 1:20000 was first referenced using the Universal Transverse Mercator (UTM) Projection. Four control points in UTM coordinates (506000,2146000; 506000,2156000; 530000,2156000 and 530000,2146000) were specified in order to calculate the transformation between digitizer and map coordinates.

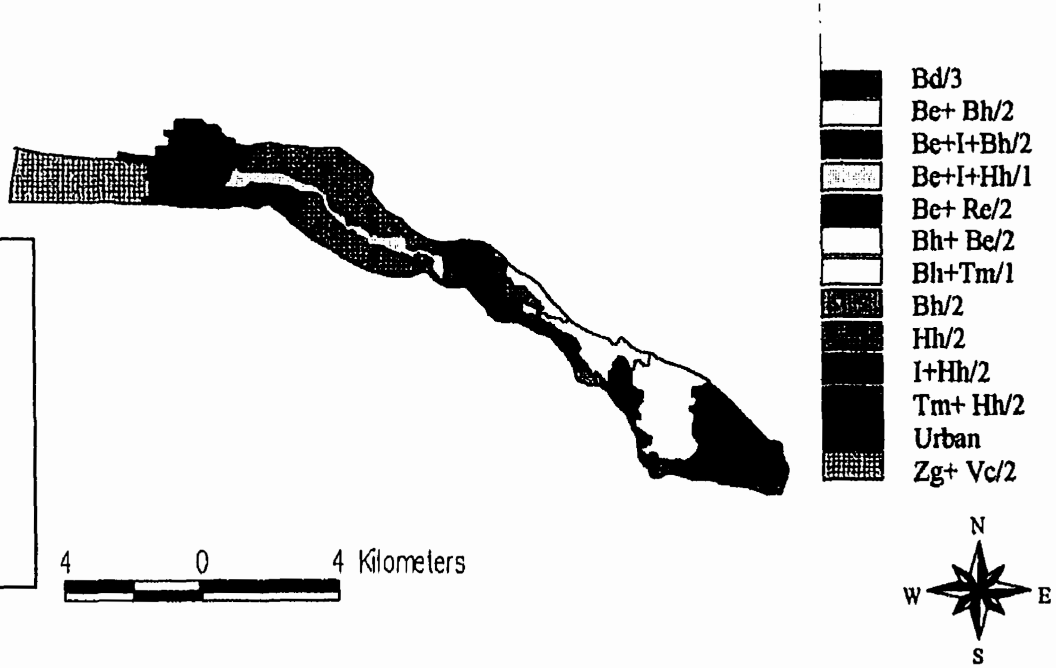
A coordinate system that included a **UTM** zone 14 for North America was used for georeferencing, and a segment (vector) map for the watershed boundary was created. The segment map was then copied and used as a base map to digitize the other cover maps so as to ensure the same scale and base map.

A soil map of **1:50000** scale published by Direccion Genral De Geografica (1983) was used to produce a soil polygon map. This soil map contains all the variables necessary (for the suitability evaluation), as attribute tables. These variables are the same variables included in the database for the proposed paradigm.

A point map shows all the samples that were taken in the watershed which was created and added to the boundary base map in order to determine the position of each sample within the watershed **(Map 2).**

Since the land suitability assessment intended as part of the objectives of this work involves not only soil parameters but climatic parameters too, maps representing the spatial variability of such climatic parameters over the study area are also part of the spatial database development. Given the fact that all of these parameters constitute point-data recorded at the discrete sites of the meteorological stations, spatial interpolation techniques were used to compute maps of the observed meteorological variables, or of variables derived from calculations using the original variables recorded. The procedures used for treatment of climatic data are given below.

Map 3: Soil polygon classes of Texcoco River Watershed



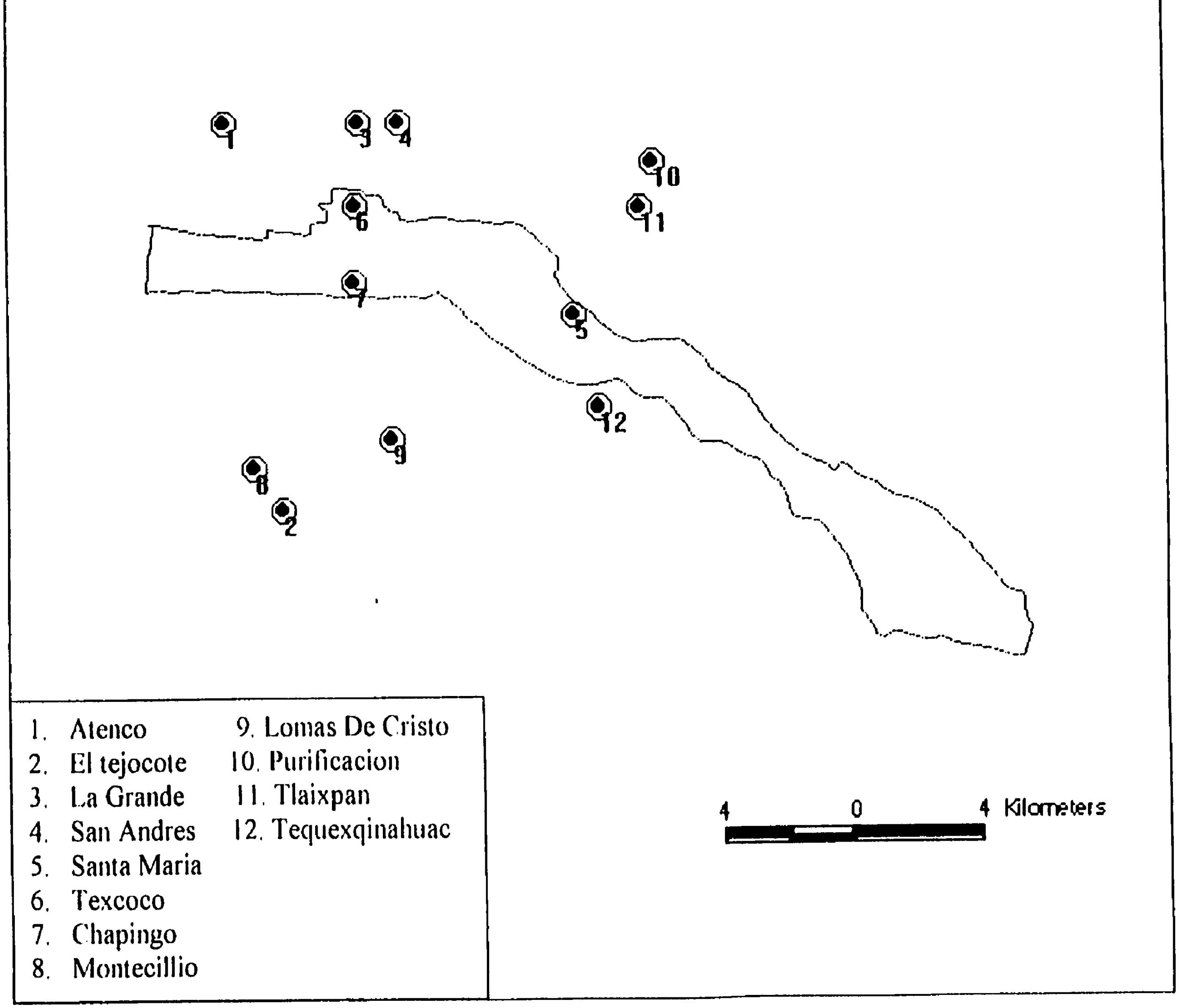
Soil Type: Bd, Be, Bh (Cambisol)
Hh (Feozem)
Re (Regosol)
Tm (Ansosol)
Zg (Solonchak)
Vc (Vertisol)
Tm (Andosol)
Soil Texture Classes: 1(Coarse), 2 (Medium) and 3(Fine)

5.6 Data Sources for Developing Land Suitability Models

The required data for building an automated land evaluation model for the study area were identified and selected from a compiled list of requirements **(section 6.7.3)** common for growing maize (FAO, 1978; Sys, 1985; Ponce-Hernandez and Beernart, 1991).

Data for 12 meteorological stations **(Map 4)** were obtained from records of the period from 1961-1988. The 12 meteorological stations selected showed some data gaps in terms of records absent for some climatic variables, namely, annual rainfall and length of growing period. Such data needed to be estimated from present records, either by interpolation/extrapolation of values, or by finding an empirical regression equation with the missing variable as dependent variable and any of the other variables with complete records as independent variables. **Appendix (2)** shows the values for minimum and mean temperature, annual rainfall, rainfall within growing season and growing period along with the coordinates for all the stations used for the analysis.

Map 4: Meteorological Stations



5.6.1 Regression Analysis for Estimation of Missing Data

In order to estimate missing data for annual rainfall and length of growing period, a regression equation was computed. This analysis set out to explore the relationships between the different climatic variables in the study area. Where a statistically significant regression equation was found with a high coefficient of determination, such equations were then used for estimation of missing data and for filling data gaps. **Map 4** shows the spatial distribution of the 12 meteorological stations in the study area.

5.6.2 Climatic data interpolation

In order to build the spatial and attribute databases for climatic requirements needed to evaluate suitability for the current and proposed paradigms, an **ASCII** format file for each variable was created and entered into the ***Surfer*** **(version 6.0)** computer program . This program provides gridding and contouring algorithms which interpolate point-data from the meteorological stations. For the purposes of interpolation and given the fact that only 5 point data were used, inverse distance functions and bicubic spline algorithms were used for interpolation of meteorological variables. This is because it was not possible to apply a Kriging technique with so a few point-data to produce a **Grid** file. The **grid** file is used to draw a plot and a contour map. Bicubic splines and inverse distance interpolation were used for the generation of a regular rectangular array of grid values contained in a **grid [GRD]** file. This method worked well in generating regularly gridded data. The smoothness of estimates derived by inverse distance methods is desirable if contouring is the final goal of estimation

(Isaaks and Strivastava, 1989). However, another rather simplistic method was used to predict the value of each climatic variable within a spatial domain, by assigning the nearest data point to each grid node. This technique creates a set of "tiles" or Thissen polygons which are spatial domains within which the value of data of the meteorological station can be extrapolated with confidence. Climatic data are commonly interpolated this way in the absence of local data from a weather station close by. The Overlay Map command in ***Surfer*** was chosen to combine the interpolated contour map resulting from use of inverse distance weighting as interpolation function. Then the maps from Thiessen polygons interpolation were produced for each variable (e.g. annual rainfall) with the soil polygon map. This polygon map represents the polygons of mapping units for the current paradigm. The overlay allows for the assigning of the values of each climatic variable to every single polygon. The values of each climatic variable for soil polygons (current paradigm) were computed by integrating the contours through a weighted average over the area of the polygon. The weights were calculated according to the distance between each two contour lines over the total area for each polygon. This is weighted mean is superior to the simple mean because it considers possible uneven spacing between contour lines. On the other hand, the soil point map contains the sample sites for point data (proposed paradigm). The soil point data map was combined with both the contour map and the Theissen polygons map for the interpolated climatic variables, in order to predict the climatic variables required for the model for each site.

5.6.3 Length of Growing Period (LGP) Data

The combination of available soil moisture and adequate temperature for crop growth is expressed in the growing period. The growing period is taken as the continuous period from the time when rainfall is greater than half the potential evapotranspiration until the time when rainfall is less than the full potential evapotranspiration, plus a number of days required to evaporate an assumed 100 mm of soil moisture reserve when available (FAO, 1978). These 100 mm of moisture are assumed since the water holding capacity of each soil is not known, thus preventing the calculation of a full water balance per meteorological station. The data used to calculate length of growing period for the study area were obtained from five meteorological stations in the watershed and its surrounding area. Historical records of evaporation and precipitation spanning about 29 years were used to calculate the potential evapotranspiration. The following empirical equation was used by Marquez Rodiles (1990) to estimate evapotranspiration **(ETP).** This equation has empirically been shown to work better in this part of Mexico than the other known equations for calculation of potential evapotranspiration. Such an equation was considered a good option given the lack of information required for other equations (e.g. Thornthwaite):

$$\mathbf{ETP = 0.8\ EV} \qquad \mathbf{(1.16)}$$

where EV is the monthly evaporation average and 0.8 is an empirical coefficient estimated for central Mexico according to the World Meteorological Organization as a substitute method for the Penman equation in situations where the radiation and wind speed terms can not be obtained, such as in the study area . The data required to calculate the **LGP** were entered into a spreadsheet for all the stations and the results were compiled and plotted in the form of climographs. These graphs were used for estimating the length of growing period.

5.7 Development of Models for Land Suitability Assessment

The Automated Land Evaluation System (**ALES**) is a computer program used for land suitability assessments. The models developed for both the current and proposed paradigms are based on the class limits of some key climatic and soil variables. Such variables define the data set required for land suitability assessment for a given crop or land utilization type **(LUT).** Matching the land quality values with land use requirements of the **LUT** is the essential part of the land evaluation exercise, which results in suitability classes. This matching process was implemented by developing and building a computer model based on decision trees for the suitability assessment of land utilization types selected within the area.

5.7.1 Definition of Land Utilization Type (LUT) for the Assessment

Maize alone was selected from the common **LUTs** and potential **LUTs** in the study area. The cultivation of maize is restricted to the rainy season between late May and early October. The successful germination of seeds and maturation of plants during other parts of the annual cycle are severely restricted by inadequate rainfall and even more so by severe winter frosts (Parson et al., 1971). As has been stated, the main objective of this study is to compare the final suitability maps produced from the two paradigms in terms of the accuracy with which suitability is predicted rather than investigate the suitability of maize in the watershed, per se.

5.7.2 Definition of Land Management Units (LMU) for the Assessment

(i) LMU for the current paradigm

The **LMUs** for the current paradigm were defined in the area according to the soil polygon delineation **(Map 3)**. The selection of the representative soil profile for each **LMU** was based on it being the central concept of the polygon. The data required for every mapping unit were obtained from a soil map scale of **1:50000** published by National Institute of Geography and Statistics (1978).

(ii) LMU for the proposed paradigm

In the case of the paradigm that is proposed here, the LMUs became the retained hard-point-data, since no polygons (mapping units) were used as generalized data. Only hard-point-data were retained, and therefore the assessment was carried out at each soil sampling locations (point-data). The required data to identify soil characteristics for each point (in this case LMU) were obtained from the soil analyses done for the 83 samples collected from the study area. However, other data such as climatic requirements were obtained from the interpolated maps of climatic parameters which were produced from ***Surfer***, then overlain on the hard-point-data representing the soil samples. In this way, both soil and climatic parameters that can be matched to the requirements of the land use can be obtained.

5.7.3 Definition of Land Use Requirements (LUR)

All the decision tree models were based on six groups of requirements. For each of the groups there is a decision tree. The decision-tree includes all the requirements within each group of limiting factors. The following are the six groups of requirements for most crops and particularly for maize (Sys, 1985):

1. Climatic requirements
2. Soil physical requirements
3. Soil fertility requirements
4. Soil salinity and alkalinity requirements
5. Soil flooding and drainage requirements
6. Topography requirements (slope).

5.7.4 Definition of Land Characteristics (LC)

Land characteristics are the measured or estimated properties of land which form the data items in the **ALES** data base. In models, they are used as the elements of the decision tree that determines the severity level of each land quality, that is the status of land quality in meeting a requirement of the crop, and ultimately the final suitability ratings of the various **LUT** by the overall assessment of individually-assessed land qualities in matching the requirements. The land characteristics chosen in this study for the land suitability evaluation are the same as those in the decision tree models. The six groups of land use requirements define the six sets of characteristics on which the land is to be evaluated and they have to be the same parameters defined in the models so that they can be matched to the **LUT** requirements during the suitability assessment. **Appendix (3) a and b** shows the total land use requirements and land characteristics considered in this study along with their defined code.

5.7.5 Suitability Classes

Land evaluation involves the assessment of the suitability of the land for a particular use. A physical evaluation is the only concern in this study, without taking into account economic considerations. The physical assessment emphasizes the relatively permanent aspects of suitability, such as climate and soil conditions, rather than changeable ones, such as prices. The land suitability classification used in this study consists of the following categories:

Two suitability orders:

S (Suitable) and **N** (Not suitable)

Six suitability classes indicating the level of limitations:

S1-0 Very suitable, no optimal limitations, optimal crop yield

S1-1 Suitable, slight limitations, almost optimal yield

S2 Moderately suitable, severe limitations, low yield

S3 Marginally suitable, severe limitations, low yield

N1 Not recommended, very severe limitations, but potentially suitable conditional upon some improvements.

N2 Not recommended, very severe limitations, unacceptable yield.

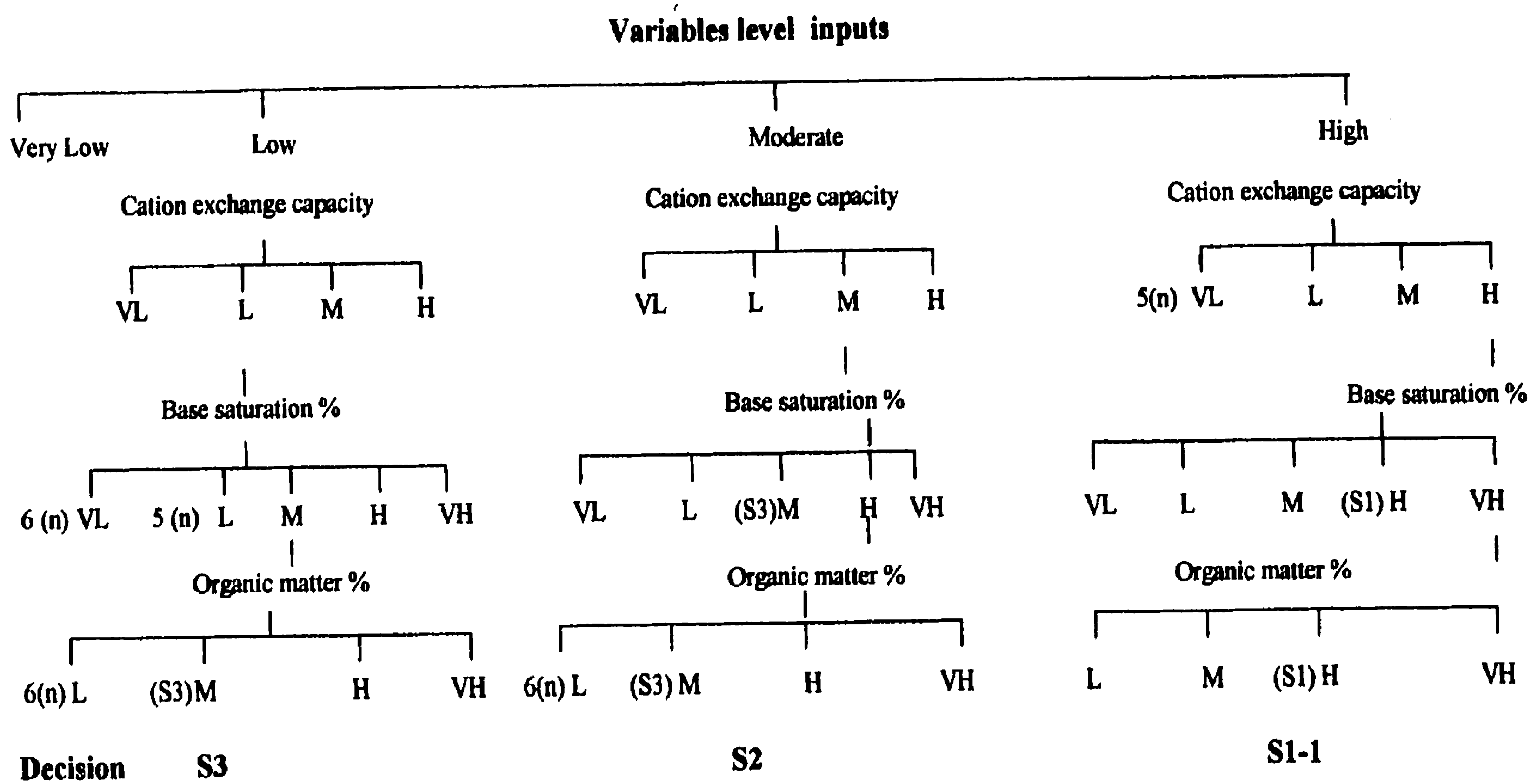
The simplest method by which **ALES** determines the physical suitability of each land unit from the set of **LUR** is the maximum limitation method. The decision about the suitability classes and sub-classes is taken by combining the effect of the different land characteristics. At the end of each branch of the tree a suitability class is assigned to summarize the suitability class of that group (e.g. climatic).

5.7.6 Models and Decision Trees

The way in which the **ALES** model builder classifies data is to build decision trees. Decision trees are hierarchical multiway keys in which the leaves are results such as land quality ratings, and the interior nodes (branch points) of the tree are decision criteria such as a land characteristic value. In the models developed for the current and proposed paradigms in this study, the physical suitability for each group of land use requirements (e.g. climatic) is assigned directly. So, instead of having the severity level for each branch the severity levels are assigned along the branches, at the interior nodes of the tree. The method developed in the model of this study is to order the land use requirements for each group according to their importance or the importance of the expected effect on crop performance. A decision is reached for one group of characteristics at a time, by following each branch in the decision tree and assigning a value for every class within the branch. Then the same process continues to the end of the tree to decide the final suitability class for this branch, as illustrated in **Fig. 6**. The levels illustrated in **Fig. 6** are shown as criteria in **Appendix 3(a)**. The process starts by choosing one level in the cation exchange capacity classes which is the most important variable in the fertility decision tree. Then the following class of input level for base saturation is chosen. All other classes remain empty at that level. This process continues to the last variable in the decision tree which is organic matter %, and the suitability class for maize production is obtained according to the maximum limitation method. The same procedure was followed for each branch and for each tree. **Appendix (4)** also shows one of the decision trees created for climatic requirements.

Data entry templates were used to specify the land characteristics for which data are to be entered into the model, and their order in the data entry form which is to be filled-in by the model user. The template is customized to the data specifically needed in the decision trees for the **LUT** (maize) in this study. On the other hand the same template is used to extract the original data from the database.

ALES provides several ways to exchange information with other computer- based systems. Database files for all land characteristics and requirements were created in a data base management system package (dBase) and exported to ALES **(Appendix 5a and b)**.



Levels of variables: VL= very low ; L= low ; M=moderate ; H= high and VH= very high
Suitability classes: 6(n)= not recommended ; 5(n)= not suitable ; S1= highly suitable ; S2= moderately suitable and S3= marginally suitable
See text and appendix for details.

Fig. 6: Soil fertility Decision Tree (the steps are taken to reach the final Decision for each branch in the decision tree)

5.8 Computation of Evaluation and Results

(i) Suitability evaluation of LUT on soil polygons: the current paradigm

The mapping units which are represented by soil polygons in the current paradigm of soil information need to be defined before performing the suitability evaluation. In this case 10 polygons were delineated to represent the study area for the current paradigm. These mapping units were included in the database files and given codes to be recognized by **ALES** during the evaluation. Once the model was fed with the data the computation took from 20 to 40 seconds for 10 polygons. This time was very short because we are evaluating only one **LUT** for all polygons.

(ii) Suitability evaluation of interpolated rasters of soils: the proposed paradigm

The procedures used for the computation of suitability under the proposed paradigm were the same steps that were taken for the current paradigm in **ALES**. However, 66 hard-point data were defined as the "management" units for the proposed paradigm. In this case every data point was evaluated separately and in this way the suitability rating at these points was predicted by the model. The land characteristics requirements used by the model were the same list for both paradigms.

5.8.1 Measurement of Maize Yield

Maize yields were measured for 48 of the 76 plots during harvest time in 1997. The method used to measure maize yields was as follow: four sub-plots of 5*2 m (10 m^2) were measured within each plot, and in each sub-plot the harvested maize cobs were weighed, followed by the rest of the above ground plant matter (stalks and leaves). These measurements are considered to represent dry weight, as the maize in Texcoco is left to dry completely in the field before harvesting. The sub-plots were located in the four corners of a square site, or in zig-zag pattern if the plot was long and narrow or terraced. The sub-plots were located several meters in from the edge of a plot to avoid edge effects. Average values from the four sub-plots were used to determine weight-per-hectare of cobs and above ground biomass for plot (Wilson, 1999). It should be noted that the measured yield might have some uncertainty due to some variations in the yield estimation procedures. These variations were induced by the timing of harvest which made it impossible to measure multiple fields simultaneously. Thus, in only a few instances yields were estimated after the farmer had harvested the field, by weighing a sample of grain sacs obtained from the whole plot and counting the number of sacs obtained.

5.8.2 Converting Suitability Classes to Yield Data for Spatial Interpolation

The land suitability classification for maize on **66** hard-point data were converted to yield data in order to interpolate the results of the assessment and generate a spatial pattern comparable to the suitability map derived from soil polygons. Each point was given a value of low input yield as described by Sys (1985) and Ponce-Hernandez and Beernart (1991). These equivalent yield values were then interpolated to produce a raster map for yield data to be used for the comparison between the yield maps resulting from the current and proposed paradigms against a set of random check sites. At each of these randomly-chosen **"check sites"** observed grain yields in the study area were recorded (the data partially courtesy of Claire Wilson, 1999 MSc thesis Watershed Ecosystem Graduate Program, Trent University).

The land suitability classes that correlated to the yield data and were used for data conversion and estimation are shown in **Table 1** (Ponce-Hernandez and Beernart, 1991).

Table 1: Land suitability classes related to yield estimation

Suitability class	Low level inputs ton/ha
S1-0	2.7-4.0
S1-1	2.25-3.6
S2	1.35-2.7
S3	0. 9-1.8
N	0-0.9

5.9 Spatial Analysis of Yield Data from Suitability Classes

The **66** hard-point-data that constitute the total data set for the proposed paradigm were included, once the suitability classes had been converted to yield values.

An **ASCII** file was created to enter the yield data points in **GEOEAS** for spatial interpolation. The **X** and **Y** variables were Eastings and Northings, respectively, whereas the **Z** variables were the soil suitability class and yield values for each data point. The analysis in the **GEOEAS** program consisted of spatial interpolation. Each variable involved was processed independently and its spatial variability was analysed by means of semi-variograms. The purpose of calculating semi-variograms is to define the nature of the spatial variation in the study area and to enable the best estimation of the semi-variogram model and its parameters needed to use the Kriging algorithm and provide Kriged estimates at previously unrecorded points.

The strategy for the computations for any given variables was as follows:

(i) Calculation of "omnidirectional" semi-variograms.

Omnidirectional semi-variograms were calculated with an angle of **90^0** to allow all pairs to be included regardless of direction, i.e. assuming isotropic variation of the variable. **Spherical** and **exponential** models were fitted to get the best estimate of the y-intercept **(nugget)**, maximum value **(sill)** and the effective range **(range)**, as well as the best type of models to be used for Kriging.

(ii) Calculation of "Directional" variograms

Semi-variograms were computed in four directions (**East-West**, **North-South**, **N - east-S west** and **N west-S east** by using an angular tolerance for each direction of **30^0**. This was done in order to determine whether there was anisotropic variation of yield over the entire area and in order to detect preferential directions of regionalization, i.e. anisotropies.

(iii) Cross-validation for choosing the best variogram model

The efficiency of the selected variogram was assessed with a cross validation procedure available in **GEOEAS**. This is a form of "Jack-knife" technique that allows for the calculation of residuals at each point left out of estimation and for which yields are known.

The available models were fitted to the experimental semi-variograms based on the criterion of greatest accuracy in predicting the yield parameters.

The residual mean square **(RMS)** between values predicted by the model and the observed values was used to assess the accuracy of prediction by each model. The RMS was calculated by "jack-knifing" at each point. These results were squared and summed (Eastman, 1997;Ponce-Hernandez, 1991). The model with smallest RMS was considered to be the most accurate for interpolation.

5.9.1 Interpolation of Yield and Grid Production

The parameters computed in the semi-variogram analysis were used for Kriging to generate grid maps for representing both suitability class and yield data. In addition to these parameters, Kriging requires specification of the type of Kriging and the search for neighbours for interpolation conditions. These conditions could be inferred from the knowledge gained from the calculated semi-variograms. Ordinary Kriging was used to perform the interpolation. A **polygon** file contained the longitude and latitude for the limits of the study area in **UTM** coordinates. This was fed to the program to draw the interpolated grid and the limits of the area. By using an in-house program the grid file was converted to an image file format and exported to the ***IDRISI** for Windows* software program to produce two raster maps of suitability and yield data for the study area.

5.10 Application of Fuzzy or Continuous Classification

The main feature of Fuzzy sets is the grouping of individuals into classes where boundaries are not, or cannot be, sharply defined. The application of Fuzzy sets in this Study was intended to create a continuous classification for the concept of **"suitable land"**, which is clearly more transitional than any crisp classification, by identifying the membership of the yield classes with different degree of membership to **"suitable land'**. The class membership was applied for the yield classes that resulted in the Kriged map. The Fuzzy membership function required control points to define the transition. These control points are used to determine the membership value at the edge of the set. From the values were given for yield estimation (**Table 1**) the membership function curve for suitable land was selected from **600 kg/ha** to **2250 kg/ha.** A yield of **600 kg/ha** has a membership of **0**, and a yield of **2250 kg/ha** has a membership of **1.0**. Between **600** and **2250**, which are the extreme points, the Fuzzy membership of yield to the concept **"suitable land"** gradually increases on the scale from **0** to **1**.

The Fuzzy module in ***IDRISI*** *for Windows* is designed for the generation of Fuzzy maps derived from ad-hoc membership functions. The module offers three types of membership functions: a sigmoidal (**"S-shaped"**), **J-shaped** and **linear function**.

A sigmoidal membership function requires four control points (**a**, **b**, **c**, and **d**) to be defined along the **x axis** (to govern the shape of the curve). In this study we use an S-shaped sigmoidal membership function, and four control points were also defined (to govern the shape of the curve) as shown in **Fig.7,** where the four control points monotonically rise but do not return to zero. The first inflection point **(a)** would be **600 kg/ha** where the suitability degree for this class starts from zero and rises at **600**, and the second **(b)** would be **2250 kg/ha**, which was assigned as the suitable land in our case. Since the curve never falls again, inflection points **c** and d would be given the same value as **b**.

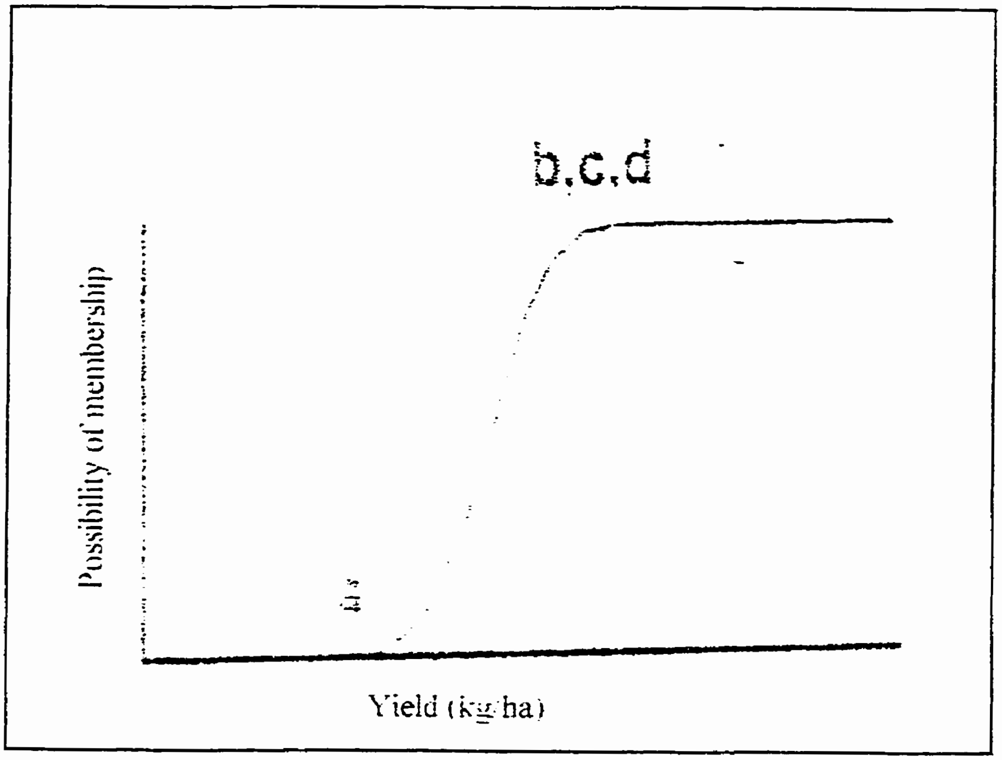


Fig. 7 Sigmoidal membership function

5.11 Comparison of Suitability Assessments Derived from the "Current" and "Proposed" Paradigms

5.11.1 Converting Fuzzy Membership Values to Yield Data

The resultant map for Fuzzy classification contains the yield data as membership values for the specific class. The Extended Cursor Inquiry option in ***IDRISI for Windows*** allows queries of the values at a specific location across the image. These values are updated for each location queried in the image. Simultaneously the three maps of Kriged, Fuzzy, and conventional classification were displayed and queried for the same locations consisting of the random check sites for which data were obtained. The following membership function was used to calculate the yield data from the membership values in the function (Burrough, 1997):

$$\mathbf{MF}_{\mathbf{A}}^{\mathbf{F}}(Z) = 1/(1 + a(Z - c)^2) \tag{1.17}$$

where MF ^{F} can be read as "the Fuzzy membership function" to the class or set **A.** The parameter a governs the shape of the function and c defines the value of the property Z, in this case yield, at the central concept. This equation was applied to determine the yield **(Z)** for the chosen location check sites from the membership values. The equation was programmed in a spreadsheet and a graph representing full membership of suitable land was created. In this study $\boldsymbol{a}$ is estimated using the central concept of the "class suitable land" as being **2250 kg/ha.** This value and the equation of the membership function allow for the calculation of the cross-over point (i.e. the "shape parameter", which is dimensionless). This parameter turned out to be **0.0000015**.

5.11.2 Comparison of the Resulting Suitability Maps

In order to judge the efficacy of the prediction of the different methods applied for the current and proposed paradigms, the estimated values of maize yield were compared to the observed maize yields at the "random check sites" by Sums of Squares of their residual from the observed as a measure of the total error in estimation. The most accurate technique was that which gave the smallest **RMS**.

These residuals were calculated with following equations:

$$\mathbf{Y_{res} = Y_{obs} - Y_{polygon}} \tag{1.18}$$

$$\mathbf{Y_{res} = Y_{obs} - Y_{kriged}} \tag{1.19}$$

$$\mathbf{Y_{res} = Y_{obs} - Y_{fuzzy}} \tag{1.20}$$

Where; $\mathbf{Y_{res}}$ = the residual value

$\mathbf{Y_{obs}}$ = the observed yield value (random check sites)

$\mathbf{Y_{polygon}}$ = the (predicted) yield value from a polygon map

$\mathbf{Y_{kriged}}$ = the (predicted) yield value from Kriged map(from hard point-data)

$\mathbf{Y_{fuzzy}}$ = the (predicted) yield value from the fuzzy map

5.11.3 Spatial Distribution of Deviations of Yield Estimates: Model Calibration

In order to provide an indication of the spatial distribution of the accuracy of yield predictions by the model, the deviations of the predicted yield values by the model from the field measured yield values at 37 random check sites were calculated and plotted on the map of the study area. Three maps of deviations of the predicted from the observed were plotted, one per each of the techniques used as predictors (i.e. polygon map, map interpolated by Kriging and the map derived from Fuzzy classification). These maps showed the areas within the watershed where the models or predictor techniques over or underestimated yield, and whether or not such deviations formed a spatial pattern. The visualization of a such pattern would allow for model calibration and for the formualtion of an explanation of model behaviour in the study area. The deviations are also expressed relative to the observed values, thus providing an indication of relative accuracy as calculated by:

$$\frac{\text{Observed} - \text{Predicted}}{\text{Observed}} * 100$$

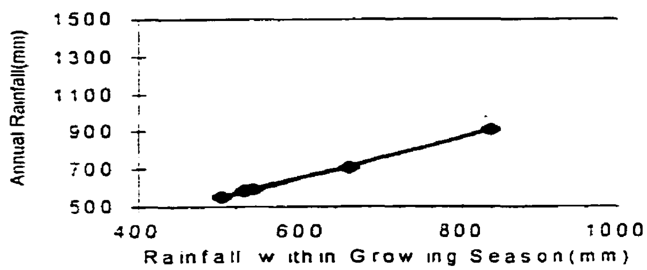
6 RESULTS

6.1 Predicting Missing Data by Regression Analysis

The results of the exploratory regression analysis, in order to develop a model to estimate missing climatic data from existing data are shown in **Fig.8 (a, b, c, d and e).**

It can be seen from these results that only the regression of annual rainfall on rainfall within growing season is significant and useful for prediction purposes (R^2 = 0.99). Hence, only this equation was used in the estimation of missing data of annual rainfall.

Regression Statistics	
Multiple R	0.999461
R Square	0.998923
Adjusted R Square	0.998564
Standard Error	5.63931
Observations	5



equation line: $y=(1.06)x+18.05$

Fig. 8 (a) Regression line of annual rainfall with rainfall within growing season

Regression Statistics	
Multiple R	0.343074
R Square	0.1177
Adjusted R Square	-0.1764
Standard Error	161.4008
Observations	5

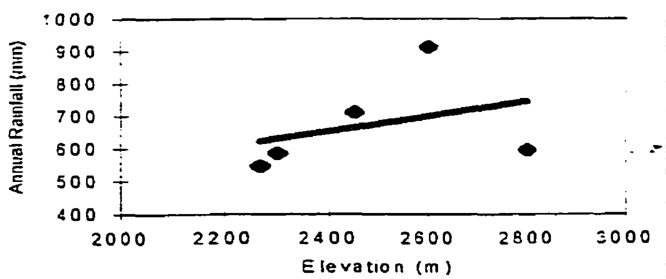


Fig. 8 (b) Regression line of annual rainfall with elevation

Regression Statistics	
Multiple R	0.28957
R Square	0.083851
Adjusted R Square	-0.22153
Standard Error	3.828627
Observations	5

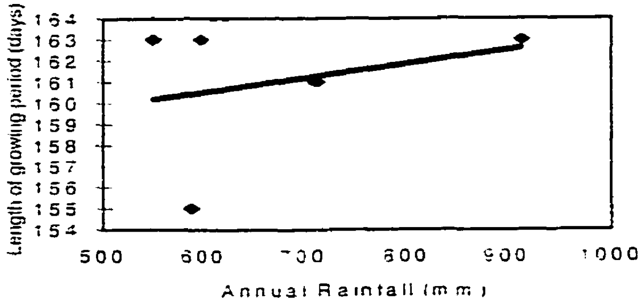


Fig. 8 (c) Regression line of length of growing period with annual rainfall

Regression Statistics	
Multiple R	0.300237
R Square	0.090142
Adjusted R Square	-0.21314
Standard Error	3.815458
Observations	5

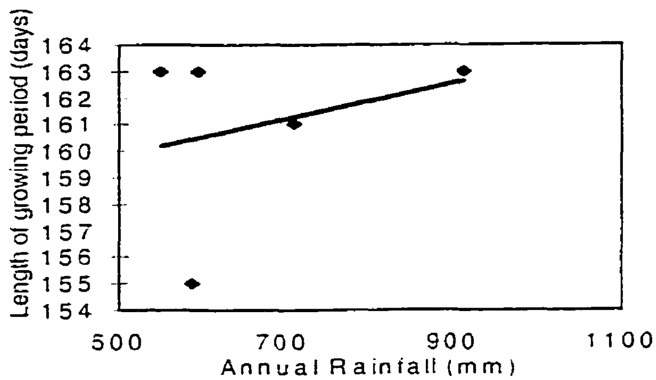


Fig. 8(d) Regression line of length of growing period with rainfall within growing season

Regression Statistics	
Multiple R	0.501997
R Square	0.252001
Adjusted R Square	0.002667
Standard Error	3.459478
Observations	5

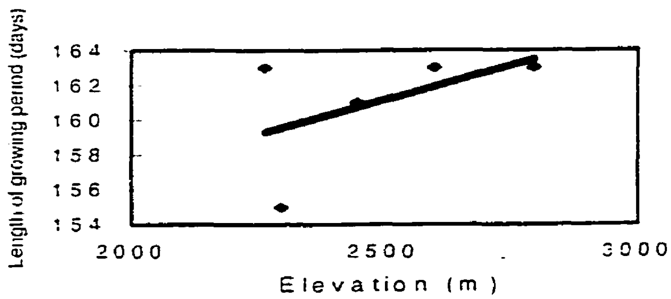


Fig. 8 (e) Regression line of length of growing period with elevation

6.2 Climatic Database Results

The climatic inventory and the methods for calculation of climatic parameters described in the preceding chapter allowed for the development of a database. **Fig. 9 (a, b, c, d** and **e)** , **Fig. 10 (a, b, c. and e), Fig. 11 (a, b, c, d, and e)** and **Fig. 12 (a, b, c, d and e)** show the resulting isoline map and Thiessen polygon maps for each variable overlain on both the soil polygon map representing the current paradigm, and on the point map for sample sites representing the proposed paradigm, over the study area. The value for each climatic requirement was then extracted by locating the point sample position within isoline interval limits for each class. However, for the current paradigm (i.e. soil polygon) a value for each climatic requirement was obtained from the isolines of interpolated maps, weighted by the respective isoline interspacing. This average value was used for all classes in the polygon. In that sense the value was generalized.

6.3 Length of Growing Period Results

Of particular interest in the climatic inventory is the calculation of the Length of Growing Period. **Fig. 13 (a, b, c, d,** and **f)** shows the climographs for the five stations used to calculate LGP and the values for each station are shown in **Table 2**.

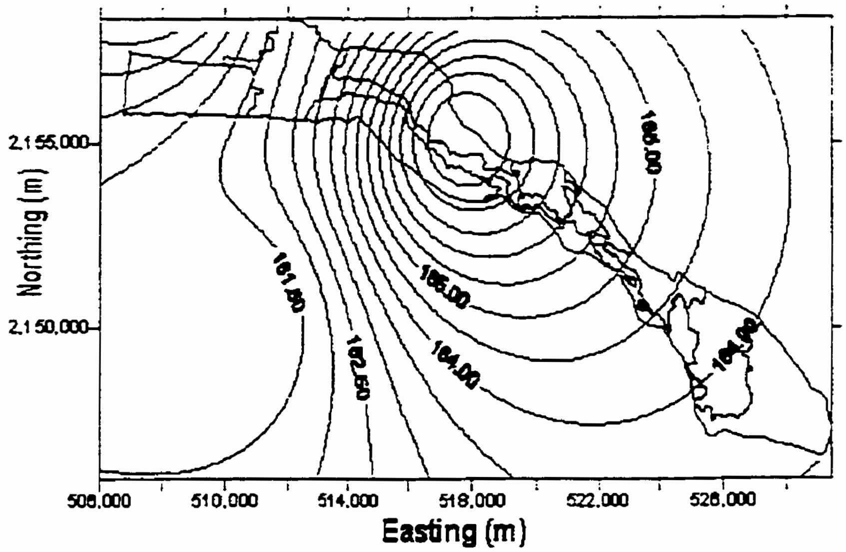

2,155,000
2,150,000
Northing (m)
508,000
510,000
514,000
518,000
522,000
526,000
Easting (m)

Fig 9(a)

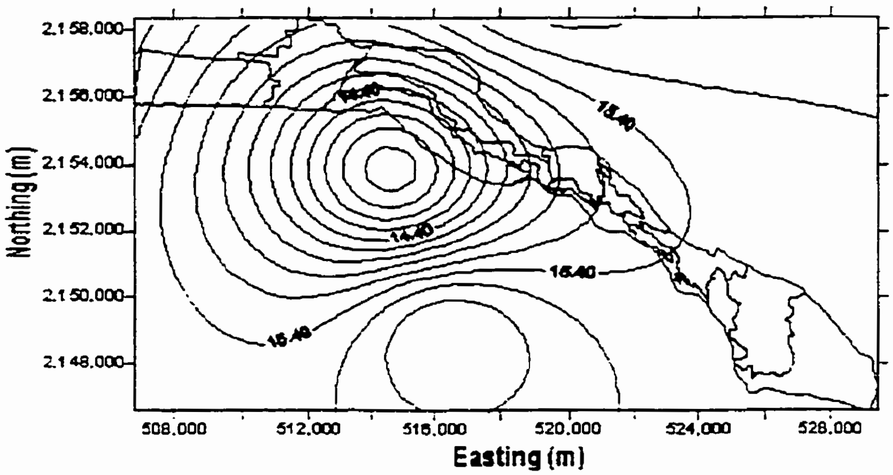

2,158,000
2,156,000
2,154,000
2,152,000
2,150,000
2,148,000
Northing (m)
508,000
512,000
516,000
520,000
524,000
528,000
Easting (m)

Fig 9(b)

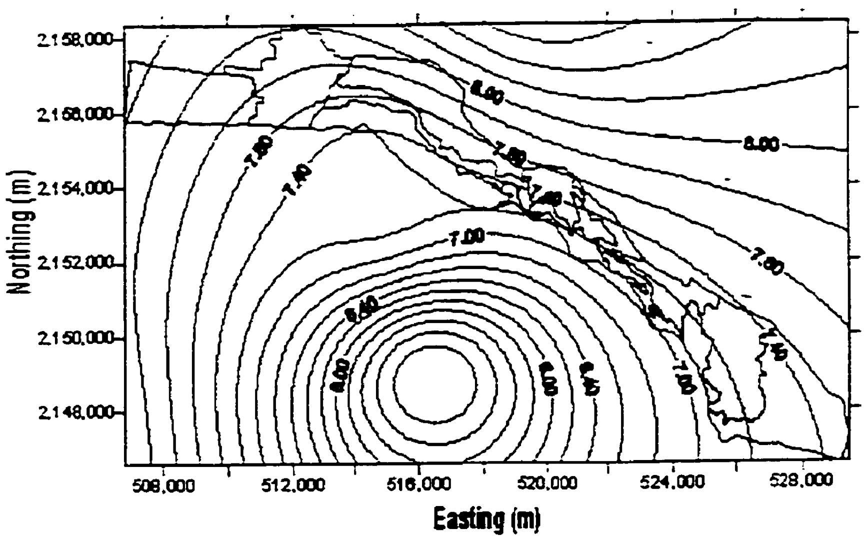

2,158,000
2,156,000
2,154,000
2,152,000
2,150,000
2,148,000
Northing (m)
508,000
512,000
516,000
520,000
524,000
528,000
Easting (m)
8.00
7.00

Fig 9(c)

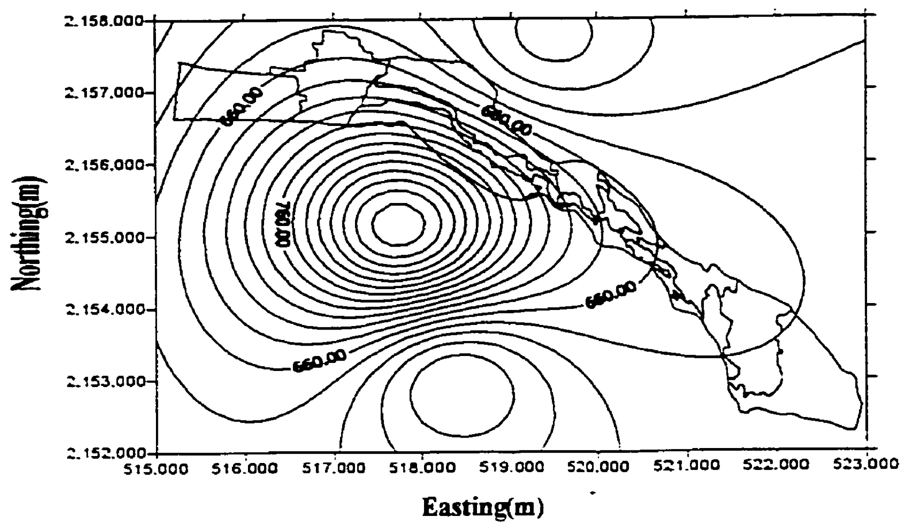

2.158.000
2.157.000
2.156.000
2.155.000
2.154.000
2.153.000
2.152.000
Northing(m)
515.000
516.000
517.000
518.000
519.000
520.000
521.000
522.000
523.000
Easting(m)
660.00
680.00
760.00

Fig 9(d)

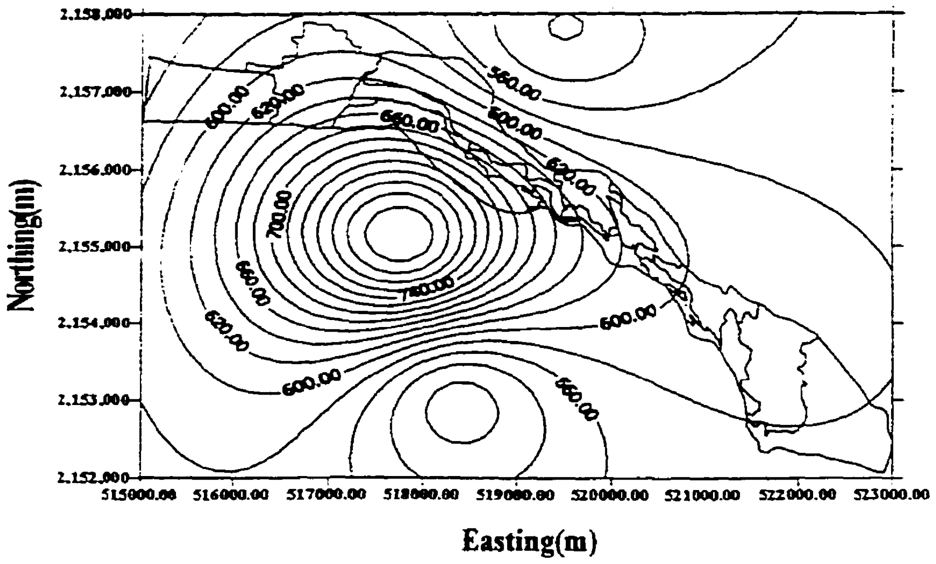


Fig 9(e)

Fig. 9 Maps resulting from overlaying the calculated isolines of climatic parameter: **a)** LGP (days); **b)** Mean temperature within growing season (^{0}C); **c)** Mean minimum temperature within growing season (^{0}C); **d)** Annual rainfall (mm) and **e)** Rainfall within growing season (mm), over the soil polygons (current paradigm).

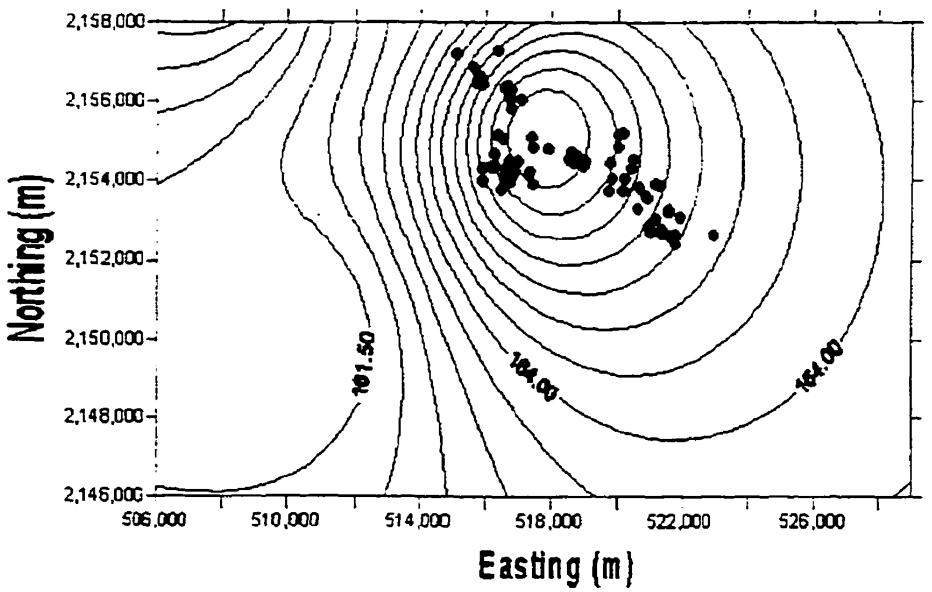


Fig 10(a)

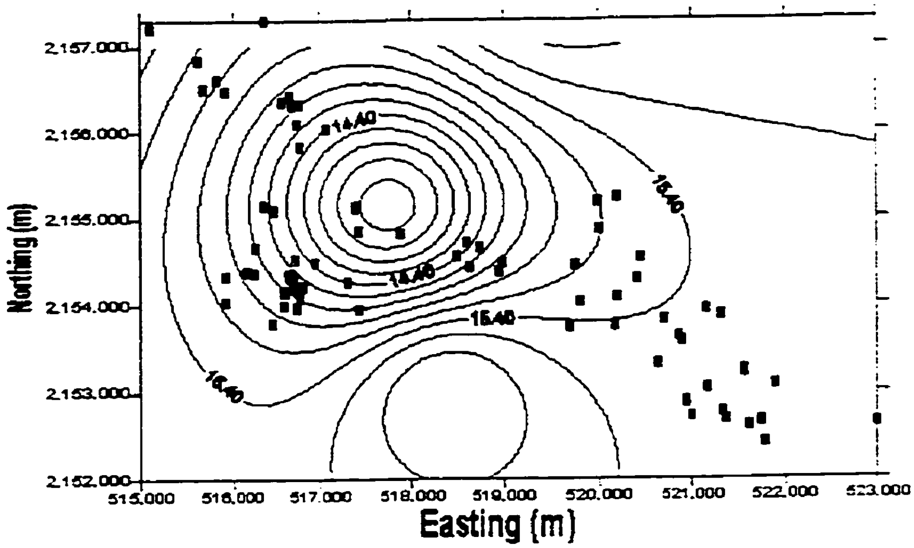


Fig 10(b)

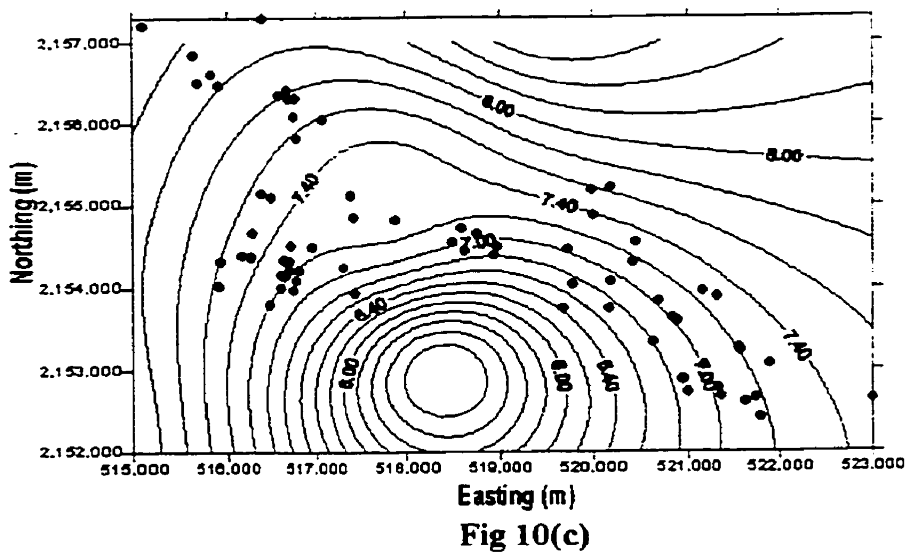


Fig 10(c)

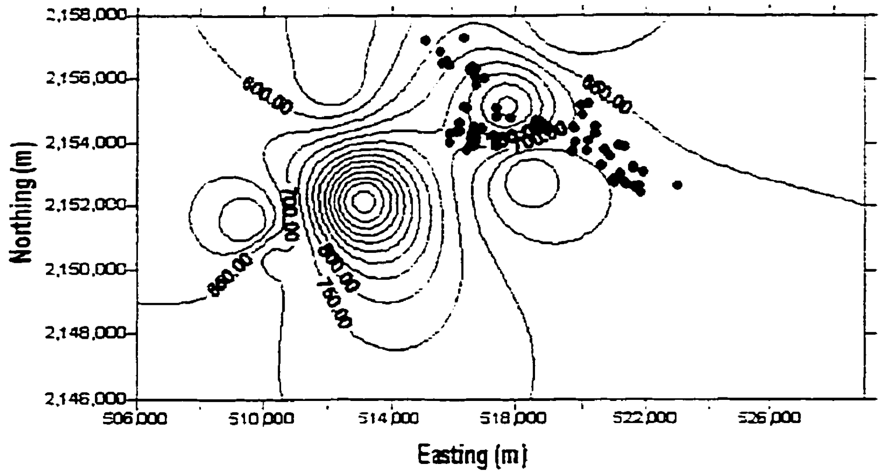


Fig 10(d)

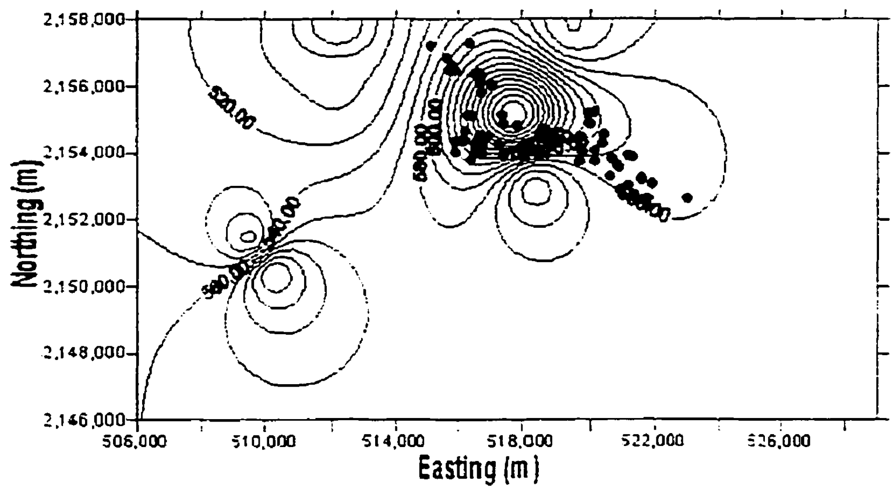


Fig 10(e)

Fig 10 Maps resulting from overlaying the calculated isolines of climatic parameter: **a)** LGP (days); **b)** Mean temperature within growing season (^{0}C); **c)** Mean minimum temperature within growing season (^{0}C); **d)** Annual rainfall (mm) and **e)** Rainfall within growing season (mm), over the point sample sites map (proposed paradigm).

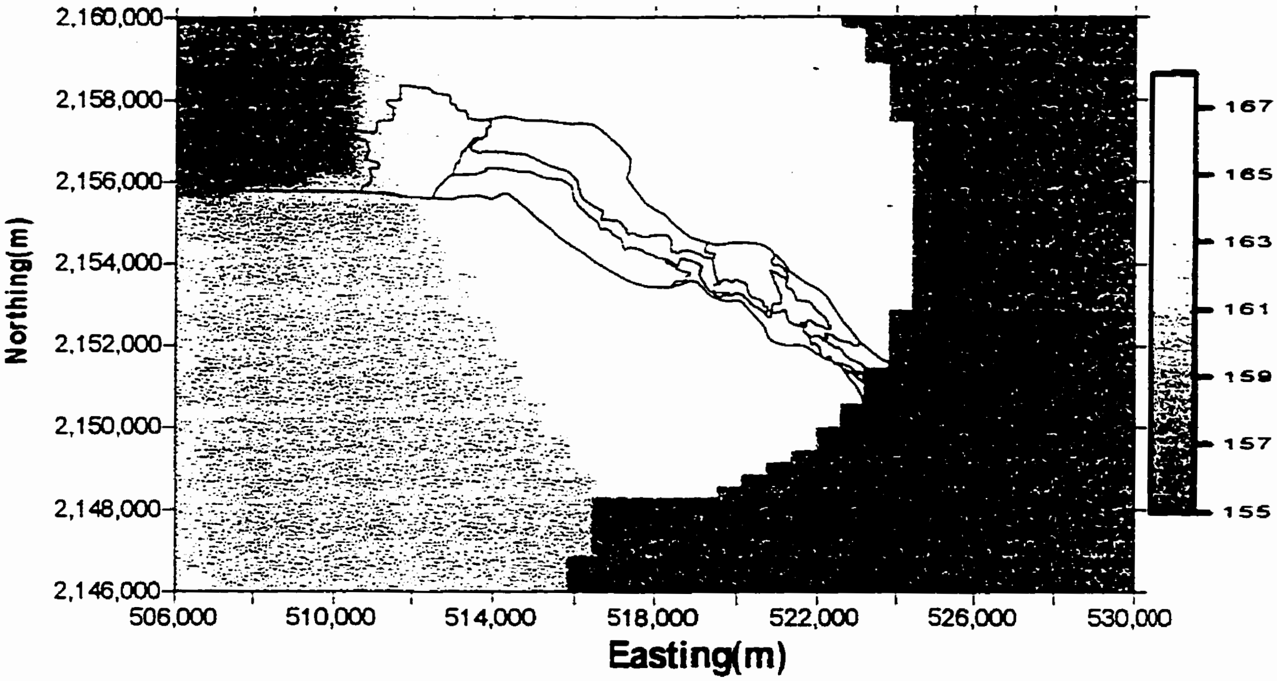


Fig 11(a)

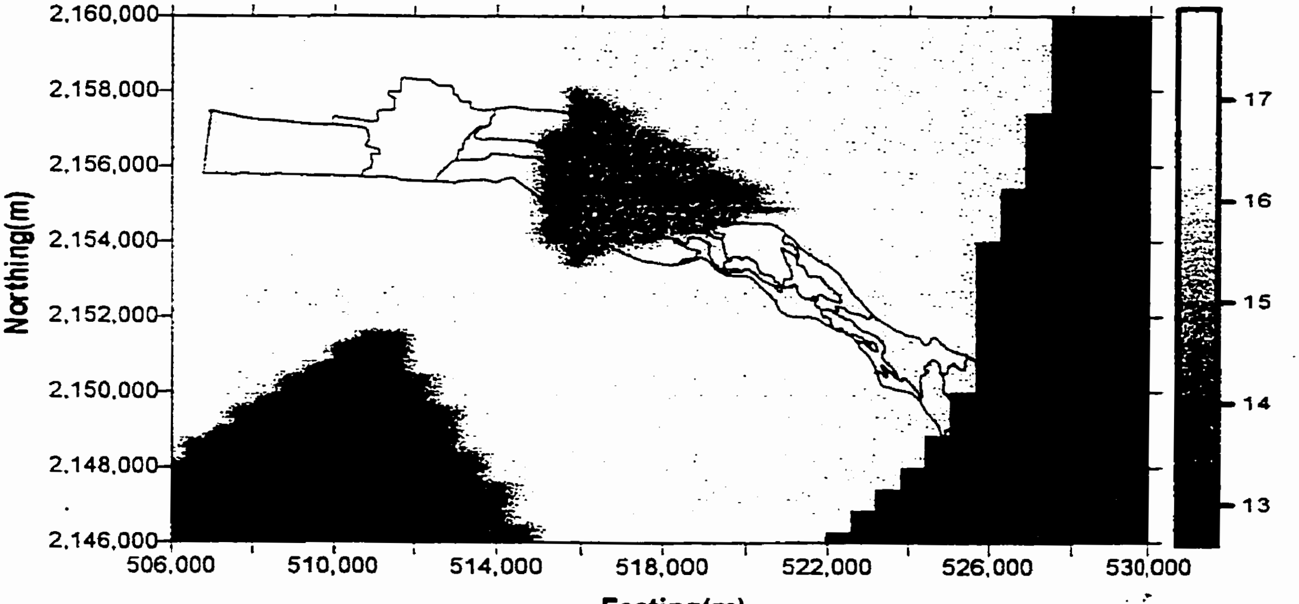


Fig 11(b)

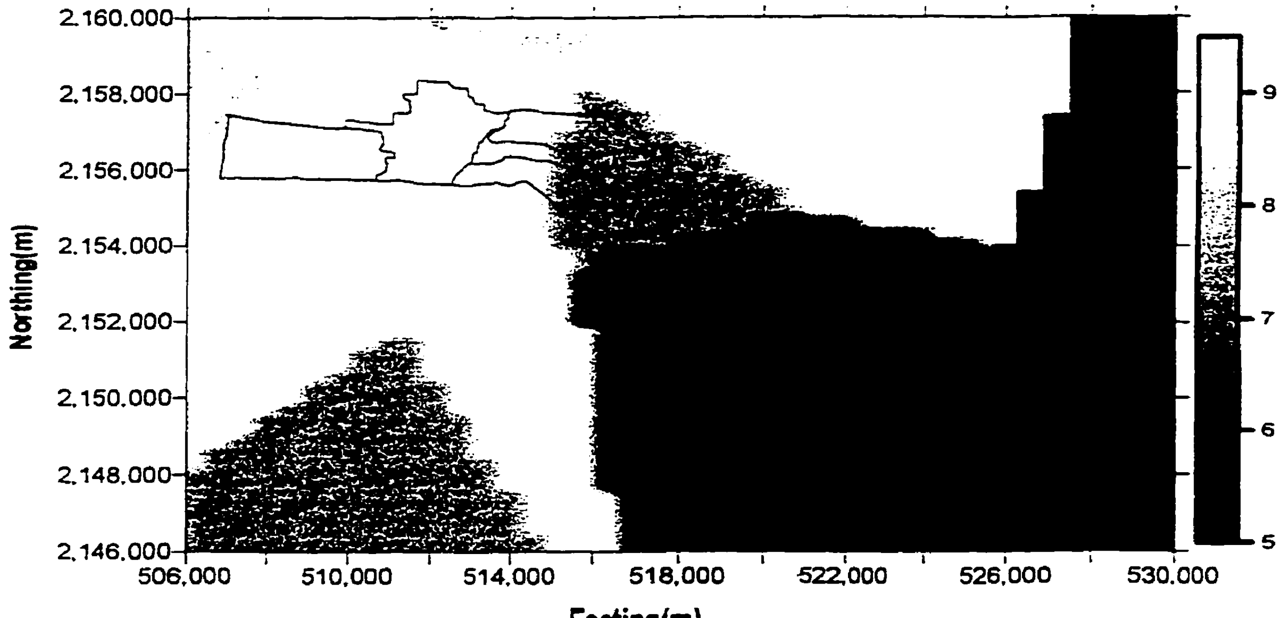


Fig 11(c)

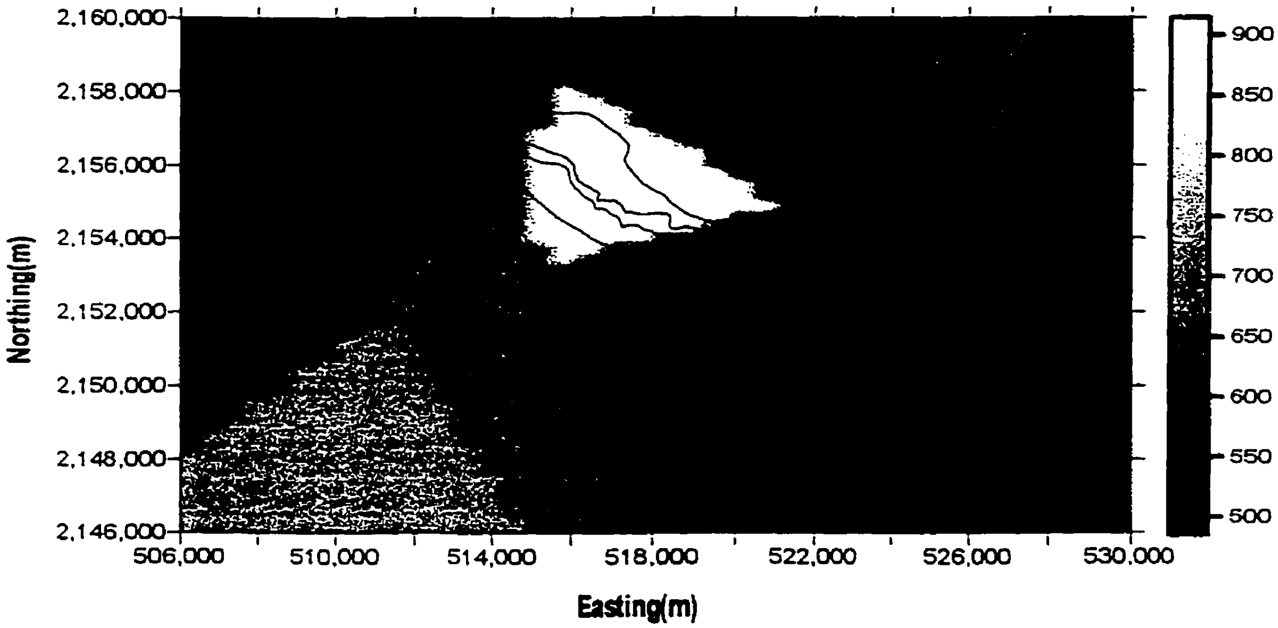


Fig 11(d)

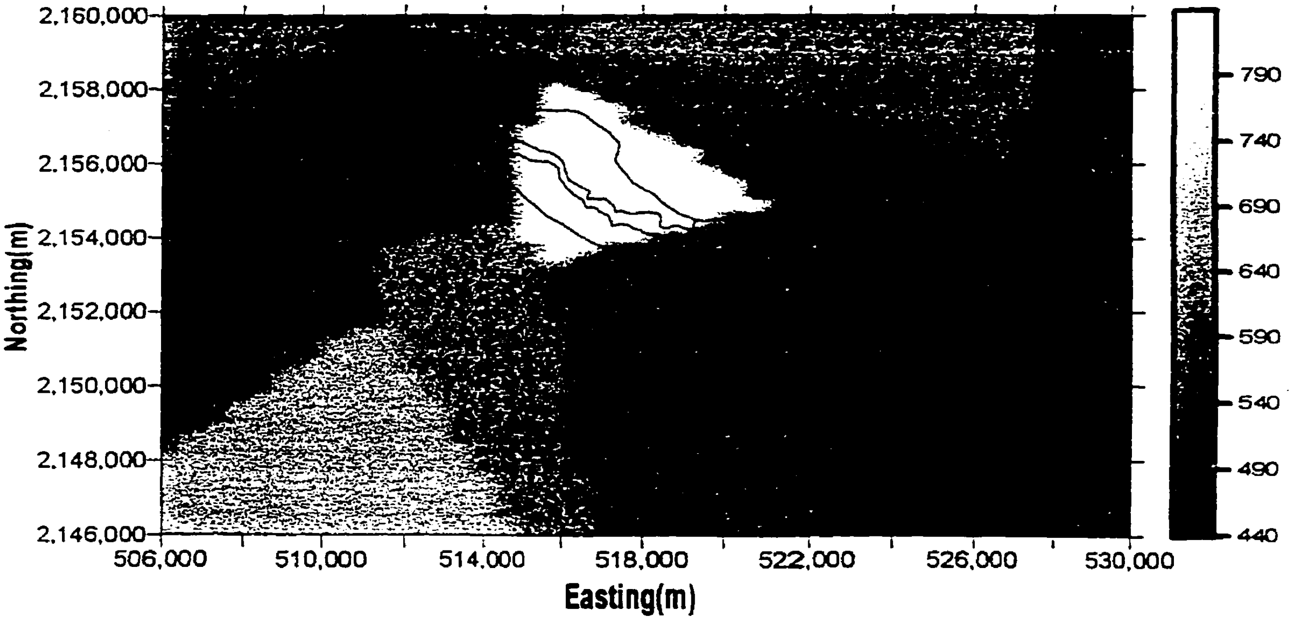


Fig 11(e)

Fig. 11 Maps resulting from overlaying the Thiessen Polygons of climatic parameter: **a)** LGP (days); **b)** Mean temperature within growing season (^{0}C); **c)** Mean minimum temperature within growing season (^{0}C); **d)** Annual rainfall (mm) and **e)** Rainfall within growing season (mm), over the soil polygons (current paradigm).

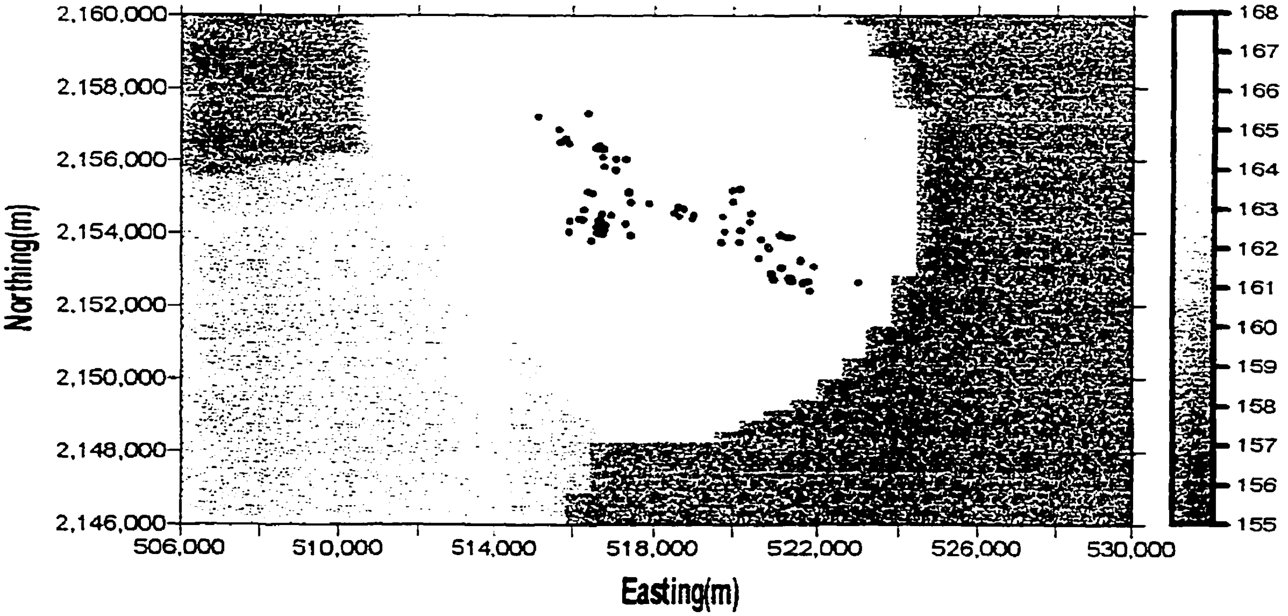


Fig 12(a)

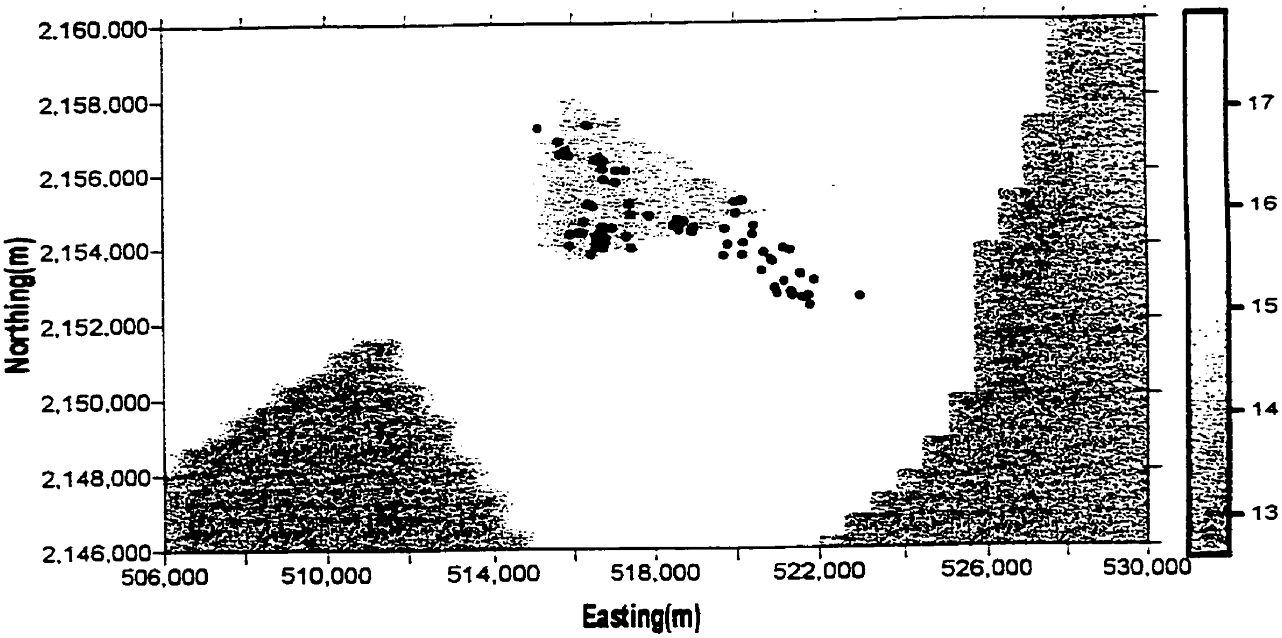


Fig 12(b)

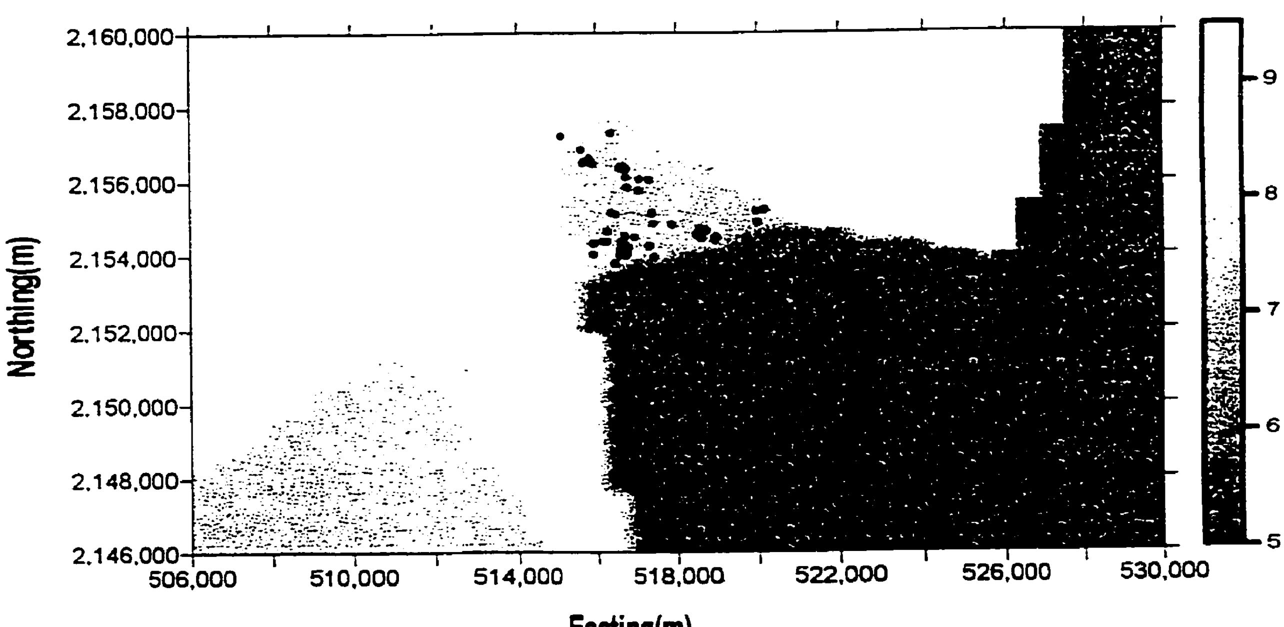


Fig 12(c)

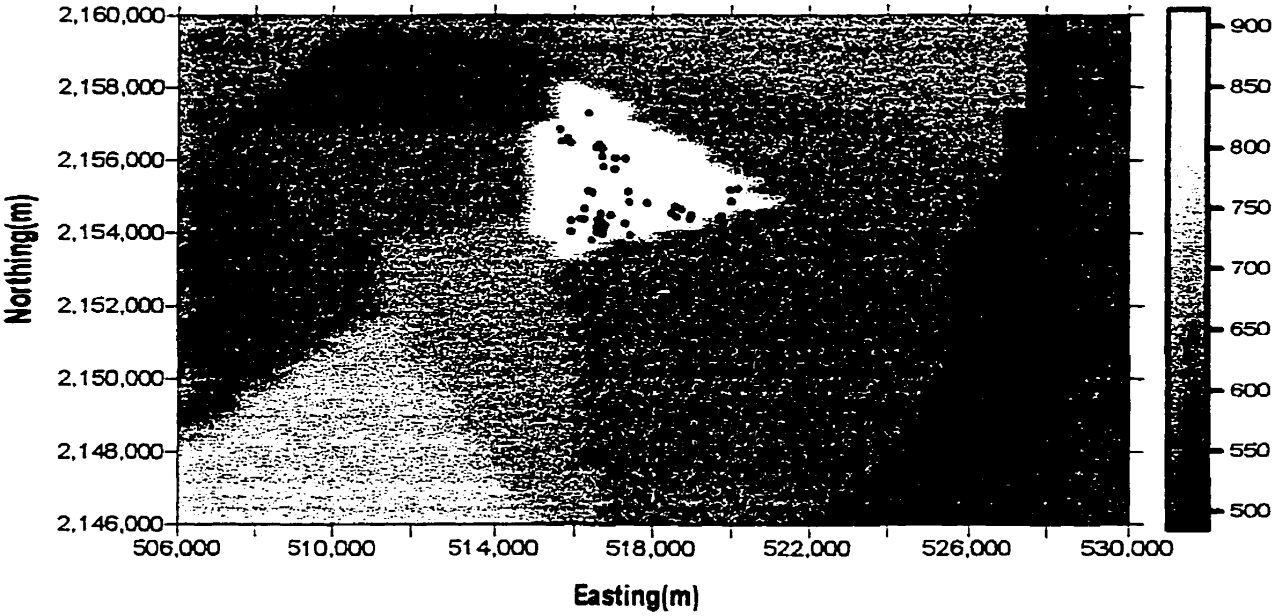


Fig 12(d)

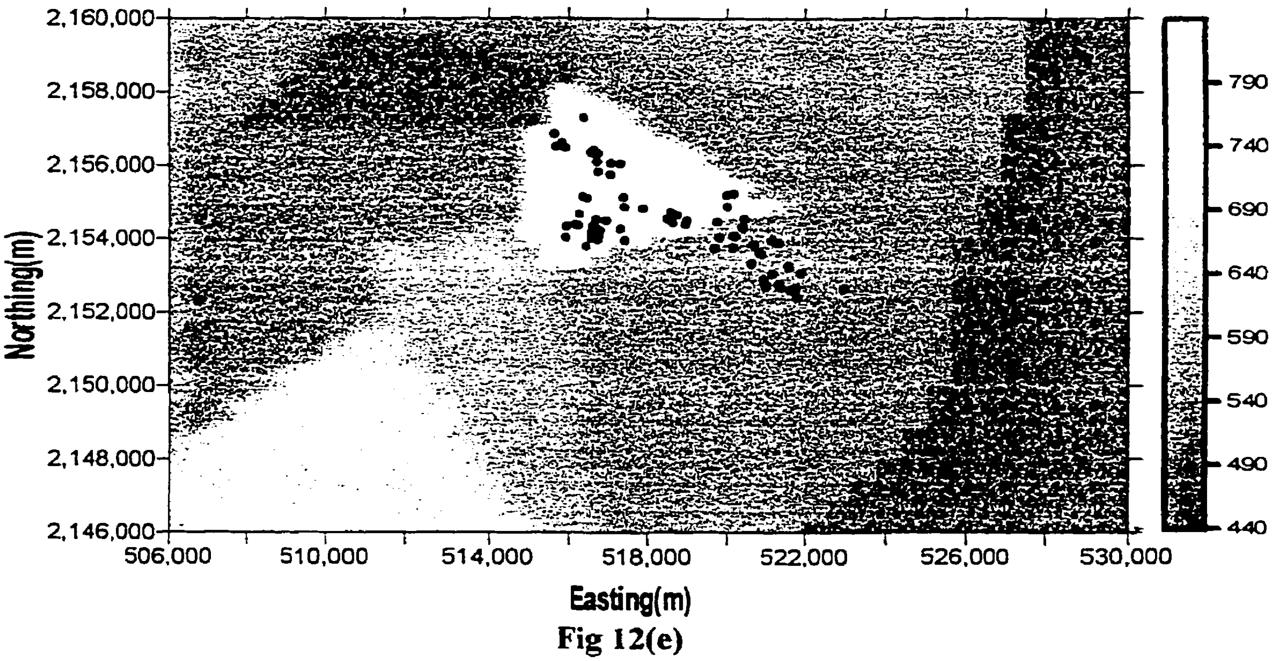


Fig 12(e)

Fig. 12 Maps resulting from overlaying the Thiessen polygon of climatic parameter: **a)** LGP (days); **b)** Mean temperature within growing season (^{0}C); **c)** Mean minimum temperature within growing season (^{0}C); **d)** Annual rainfall (mm) and **e)** Rainfall within growing season (mm), over the point sample sites map (proposed paradigm).

Fig. 13 Length of Growing Period for five stations in Texcoco area

p- precipitation ETP- evapotranspiration a- beginning of growing period

b- end of growing period

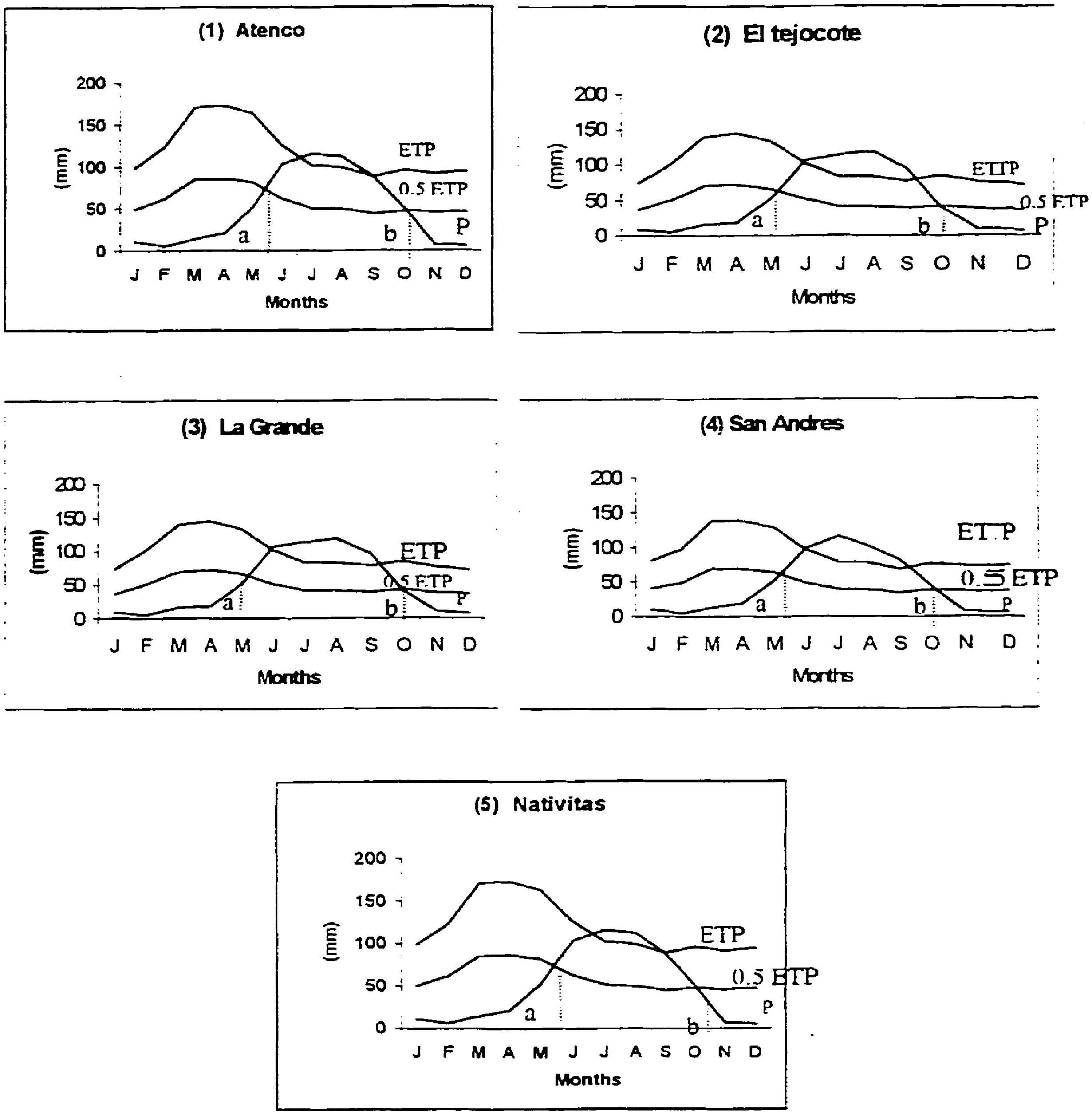


Table 2: Length of growing period values (days) calculated for the 5 meteorological stations.

Code	Station Name	LGP Calculated
1	Atenco	155
2	El tejocote	161
3	La Grande	163
4	San Andres	163
5	Nativitas	168

6.4 The Current Paradigm of Soil Information: evaluation results

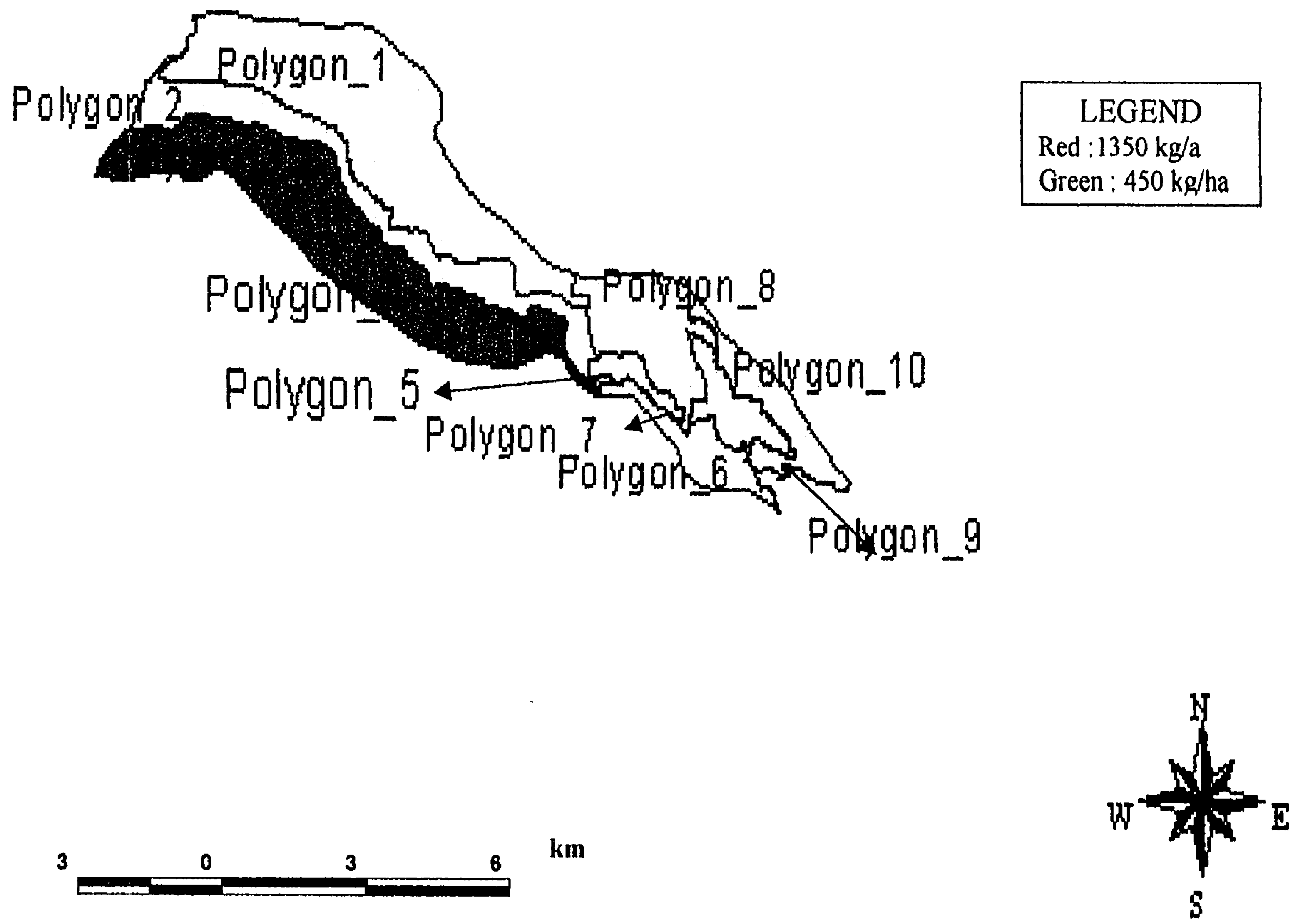
The evaluation results matrix is a two-dimensional array with rows being the map units that were evaluated (here, polygon1, polygon2...polygon10), and the columns being the land utilization types for which an evaluation was computed. In this case only maize was evaluated, and the corresponding suitability class is shown in each cell. The results of the evaluation for the current paradigm **(Table 3)**, in the form of suitability matrices, were imported into the GIS component of the Integrated Land and Watershed Information System (ILWIS). Using the creating module in (ILWIS) the segments map of the study area for soil polygons **(Map 3)** were polygonized **(Map 5)**. Attribute files were created with them and labels were assigned for polygons of an already digitized soil map. The raster suitability map **(Map 6)** that resulted will be compared to the map that will be produced for the proposed paradigm.

Table 3 The suitability matrix for the current paradigm

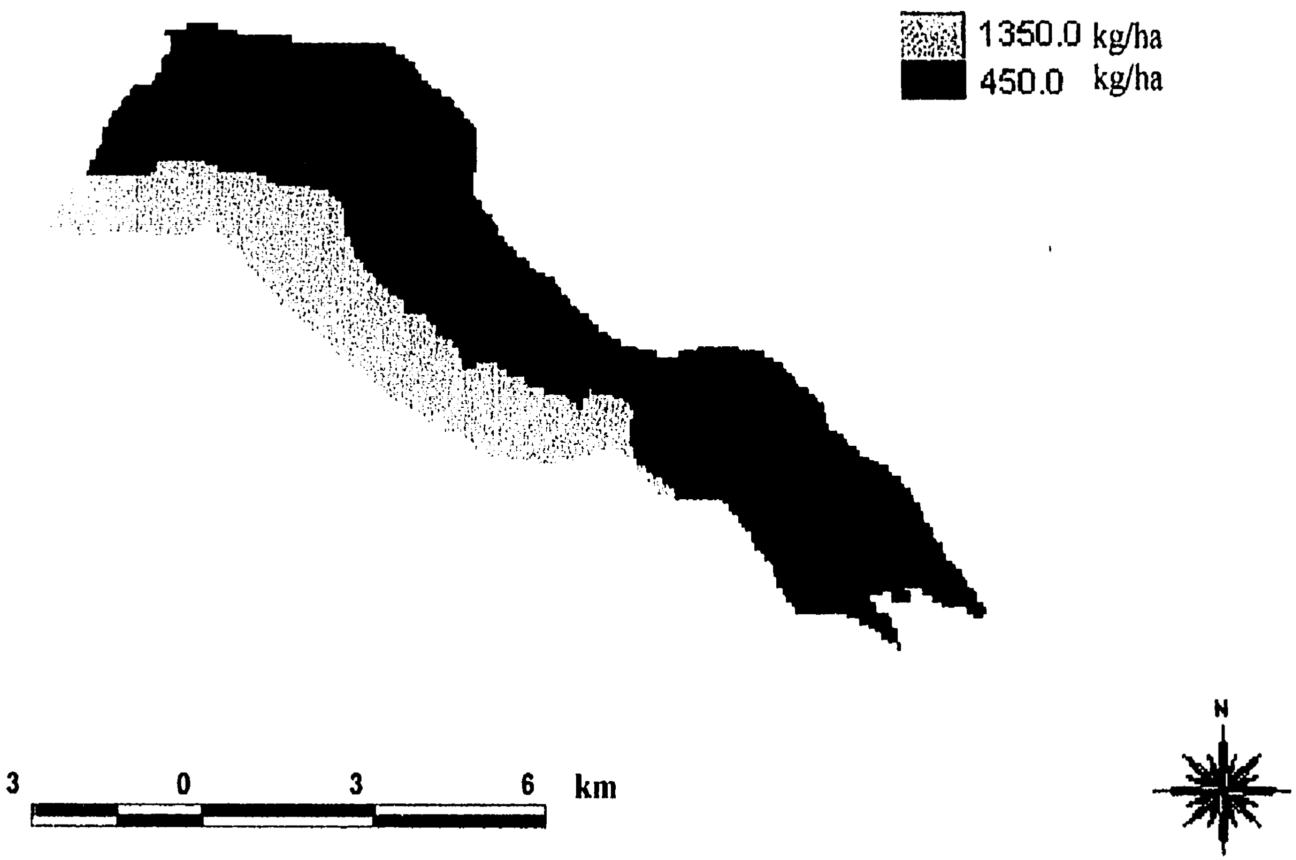
Polygon ID	Land Utilization Type (LUT)	Resultant Suitability Class	Limiting Factor
"1"	"MA"(Maize)	",6,"	,"6S"
"2"	"MA"(Maize)	",6,"	,"6S"
"3"	"MA"(Maize)	",4,"	,"4N/T/c"
"4"	"MA"(Maize)	",4,"	,"4N/T/c"
"5"	"MA"(Maize)	",4,"	,"4N/T/c"
"6"	"MA"(Maize)	",6,"	,"6S"
"7"	"MA"(Maize)	",6,"	,"6S"
"8"	"MA"(Maize)	",6,"	,"6S"
"9"	"MA"(Maize)	",6,"	,"6S"
"10"	"MA"(Maize)	",6,"	,"6S"

- Limiting factors: (N: Salinity, S: Soil physical characteristics, c: Climatic and T: Topography
- Numbers before the limiting factors indicate the suitability classes (3= S2, 4 =S3 and 5=N)

Map 5: Predicted yield by suitability classes on soil mapping units or polygons (current paradigm)



Map 6: Predicted yield by suitability classes on soil polygons or mapping units (current paradigm): Rasterization of map (5)



6.5 The Proposed Paradigm of Soil Information

The same approach to evaluating the accuracy of suitability assessments for maps derived from soil polygons, part of the so called "current paradigm", was also adopted for sample point data (proposed paradigm). As a part of the proposed paradigm, methods based on spatial interpolation and fuzzy set theory were used. The results of applying such techniques to the sample point data (the proposed paradigm) are shown in **Table 4**. Suitability matrices were calculated. These matrices predict the suitability rating for targeted points on the area. These ratings at target points were then converted into crop yield equivalents in order to be able to estimate from spatial interpolation the suitability at any point within the study area.

Table 4 Land suitability evaluation matrix for hard-point data (proposed paradigm)

Site ID	Land Utilization Type (LUT)	Resultant Suitability class	Limiting factor
"1"	"MA" (Maize)	",5,"	"5N"
"2"	"MA" (Maize)	",4,"	"4/T/c"
"3"	"MA" (Maize)	",4,"	"4c"
"4"	"MA" (Maize)	",4,"	"4c"
"5"	"MA" (Maize)	",4,"	"4c"
"6"	"MA" (Maize)	",4,"	"4/T/c"
"7"	"MA" (Maize)	",4,"	"4/T/c"
"8"	"MA" (Maize)	",4,"	"4/T/c"
"9"	"MA" (Maize)	",4,"	"4/T/c"
"10"	"MA" (Maize)	",4,"	"4c"
"11"	"MA" (Maize)	",4,"	"4c"
"12"	"MA" (Maize)	",5,"	"5c"
"13"	"MA" (Maize)	",5,"	"5c"
"14"	"MA" (Maize)	",4,"	"5c"
"15"	"MA" (Maize)	",4,"	"4c"
"16"	"MA" (Maize)	",4,"	"4/T/c/f"
"17"	"MA" (Maize)	",4,"	"4/T/c/f"
"18"	"MA" (Maize)	",4,"	"4/T/c/f"
"19"	"MA" (Maize)	",4,"	"4/T/c/f"
"20"	"MA" (Maize)	",4,"	"4/c/f"
"21"	"MA" (Maize)	",4,"	"4/c/f"
"22"	"MA" (Maize)	",4,"	"4/c/f"
"23"	"MA" (Maize)	",4,"	"4/c"
"24"	"MA" (Maize)	",4,"	"4/c/f"
"25"	"MA" (Maize)	",4,"	"4/c/f"
"26"	"MA" (Maize)	",4,"	"4/c/f"
"27"	"MA" (Maize)	",4,"	"4/c/f"
"28"	"MA" (Maize)	",4,"	"4/f"
"29"	"MA" (Maize)	"4,"	"4/c/f"
"30"	"MA" (Maize)	",4,"	"4/c/f"
"31"	"MA" (Maize)	",4,"	"4/c/f"
"32"	"MA" (Maize)	",4,"	"4/c/f"
"33"	"MA" (Maize)	",4,"	"4/T/c/f"
"34"	"MA" (Maize)	",4,"	"4/T/c"
"35"	"MA" (Maize)	",5,"	"5c"
"36"	"MA" (Maize)	",4,"	"4/c/f"
"37"	"MA" (Maize)	",4,"	"4/T/c/f"
"38"	"MA" (Maize)	",4,"	"4/c/f"
"39"	"MA" (Maize)	",4,"	"4/c/f"
"40"	"MA" (Maize)	",4,"	"3/c"
"41"	"MA" (Maize)	",4,"	"4/c"
"42"	"MA" (Maize)	",3,"	"3/c/f"

Cont'd.

"43"	"MA" (Maize)	",3,"	"3/c/f"
"44"	"MA" (Maize)	",3,"	"4/T/c/f"
"45"	"MA" (Maize)	",4,"	"4/c"
"46"	"MA" (Maize)	",4,"	"4/f"
"47"	"MA" (Maize)	",4,"	"4/c/f"
"48"	"MA" (Maize)	",4,"	"4/c/f"
"49"	"MA" (Maize)	",4,"	"4/c/f"
"50"	"MA" (Maize)	",4,"	"4/T/c/f"
"51"	"MA" (Maize)	",4,"	"4/T/c/f"
"52"	"MA" (Maize)	",4,"	"4c/f"
"53"	"MA" (Maize)	",4,"	"4c"
"54"	"MA" (Maize)	",4,"	"4/c/f"
"55"	"MA" (Maize)	",4,"	"4/c/f"
"56"	"MA" (Maize)	",4,"	"5/c"
"57"	"MA" (Maize)	",5,"	"5/c"
"58"	"MA" (Maize)	",5,"	"5/c"
"59"	"MA" (Maize)	",5,"	"4/c/f"
"60"	"MA" (Maize)	",4,"	"4/c/f"
"61"	"MA" (Maize)	",4,"	"4/c/f"
"62"	"MA" (Maize)	",4,"	"4/c/f"
"63"	"MA" (Maize)	",4,"	"5/c"
"64"	"MA" (Maize)	",5,"	"5/c"
"65"	"MA" (Maize)	",5,"	"4S/T/c/f"
"66"	"MA" (Maize)	",4,"	"4T/c/f"

* Limiting factors: (N: Salinity, S: Soil physical characteristics, c: Climatic;T: Topography) and f: fertility

Numbers before the limiting factors indicate the suitability classes (3= S2, 4 =S3 and 5=N)

6.5.1 Spatial Analysis

Summary statistics for 66 sampling data points are given in **Table 5**. The principal tool in the use of **Kriging** interpolation is the experimental semi-variogram analysis. First, the distance class intervals (lag) and directional tolerances for computing the experimental semivariograms of yield point data from conversion of suitability ratings were specified. The usual practice is to compute and plot variograms along transects in several directions and compare them visually (Burgess and Webster, 1980). The **omnidirectional** variogram **(Fig. 14)** with **90** ^{0} tolerance shows a well-defined spatial structure except for the fifth point, which is low. The default lag intervals are computed from a rule-of thumb which states that variograms are generally not valid beyond one-half the maximum pair distance (Englund and Sparks, 1991). Therefore the maximum pair distance was divided by two, and then subdivided into ten equal distance classes. Different lag distances were traded-off to include the maximum number of pairs. This causes an increase in noise comparable to the default lag, which is **455m**.

Following the estimation of the lag distance, the of Regionalized Variable Theory called for the fitting of one of several of models provided by the **GEOEAS** program used in the calculation. **Gaussian** and **Exponential (Fig. 15 and 16)** models were fit to the omnidirectional semivariogram and both of the two models proposed were satisfactory. The parameters of the isotropic model (omnidirectional) selected are shown in **Table 6**. However, it should be noted that the sill value in the program **GEOEAS** is calculated by removing the nugget effect from the total value of the sill. Upon completion of the omnidirectional semivariogram modelling the directional nature of variability was explored. The semivariograms were fit using **"true"** directional tolerance of **30**^{0} for four directions at angles of **0, 45, 90,** and **135** degrees. These four directions are likely to show an erratic and irregular behaviour especially in the **East-West** and **N-west-S east** directions and this can be explained by the short distance between the width of the watershed in these directions, which tend to be very narrow. The directions **North-South** and **N-east-S-west (Fig. 17** and **18)** have a break in the semivariogram around the **2000** and **4000** m distances. This discontinuity tends to confirm the range of influence in the omnidirectional models where the semivariograms have the same breaks at the same distances. This is presumably due to the shape of the watershed and the change in the temperature from the north to the south direction. The selection between two candidate models fitted to directional data is considered an important part in the process of spatial interpolation. The accuracy of the semivariogram models was checked using a cross-validation technique in the **x-valid** section of the **GEOEAS** program. This amounts to applying a "Jack-knife" technique. The Residual Mean Squared **(RMS)** values for the models compared after cross validation were **0.63** for the **Gaussian** model and **0.93** for the **exponential** model. After the estimates of RMS for the semivariograms were compared, the Gaussian model was fitted and the parameters of the model were used for spatial interpolation.

Table 3: Batch statistics for suitability classes and yield data obtained from STATS menu in GEOEAS

	Suitability Classes	Yield kg/ha
N of points used	66	66
N of points missing	0	0
Mean	4.106	1712.9
Variance	0.189	151634.3
Std. Dev.	0.434	389.40
Coef. Var.	10.576	22.92
Skewness	0.572	-0.572
Kurtosis	4.780	1.9
Minimum	3.0	900
Median	4.0	1800
Maximum	5.0	2700

Table 4: Isotropic models parameters

Model	Nugget	Sill	Range (m)
Gaussian	49700	164400	1110
Exponential	16700	166000	1455

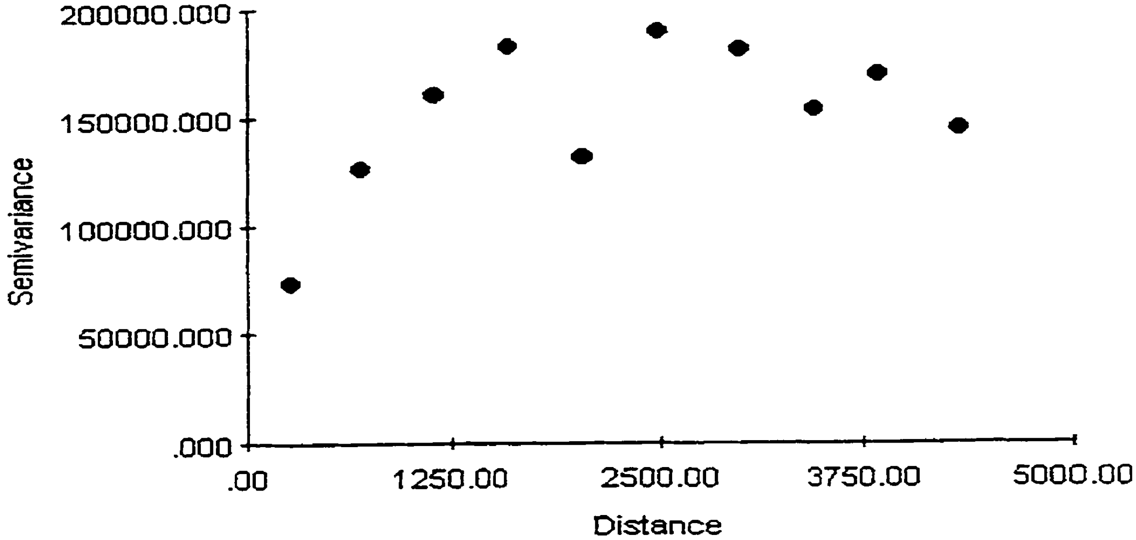


Fig.14 Omidirectional variogram for yield

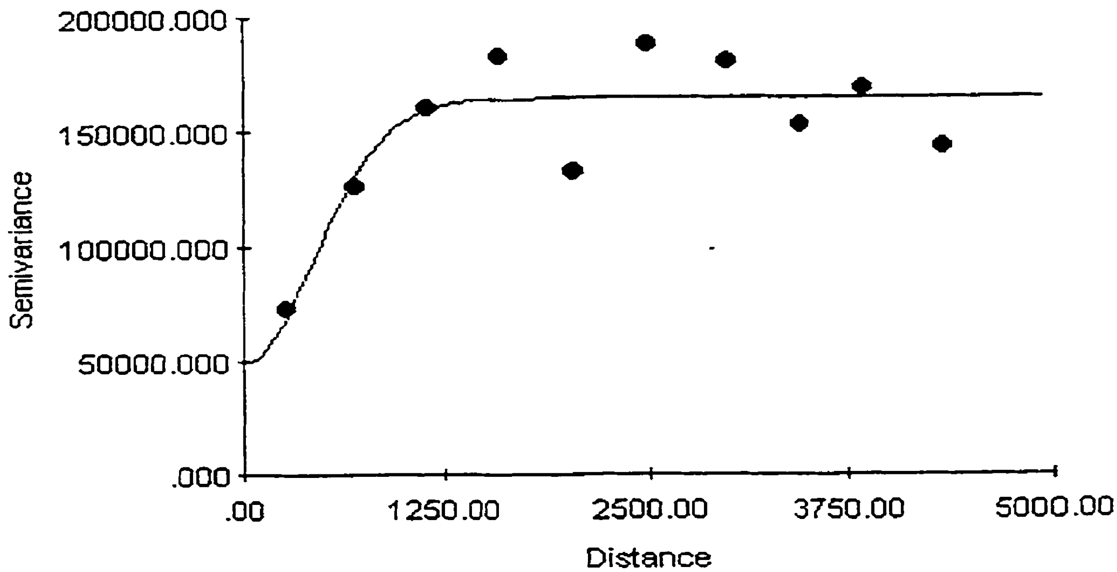


Fig. 15 Variogram for yield Texcoco Watershed;
Gaussian model

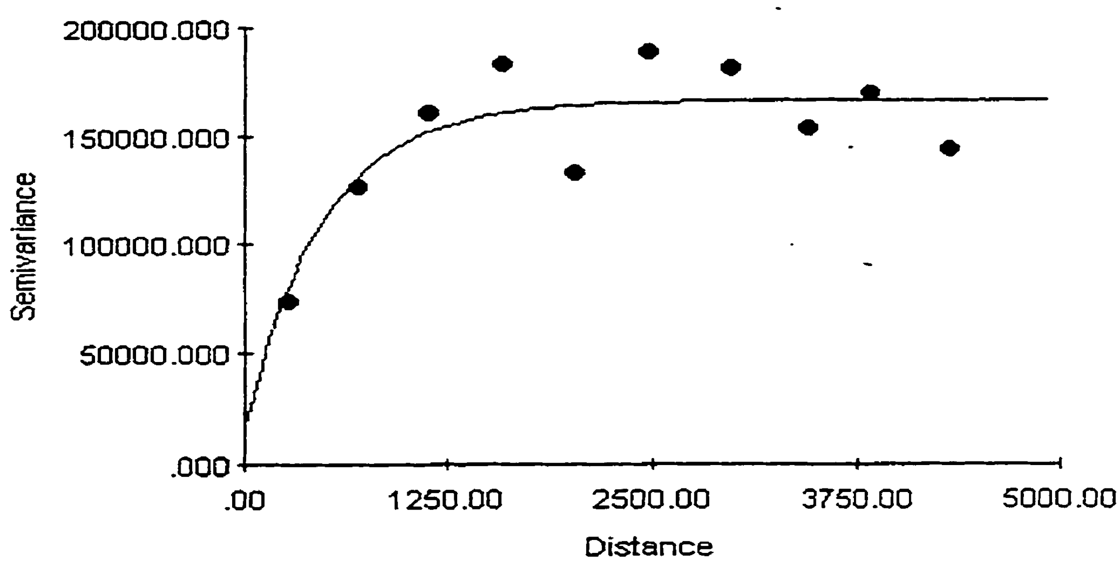


Fig. 16 Variogram for yield Texcoco Watershed: Exponential model

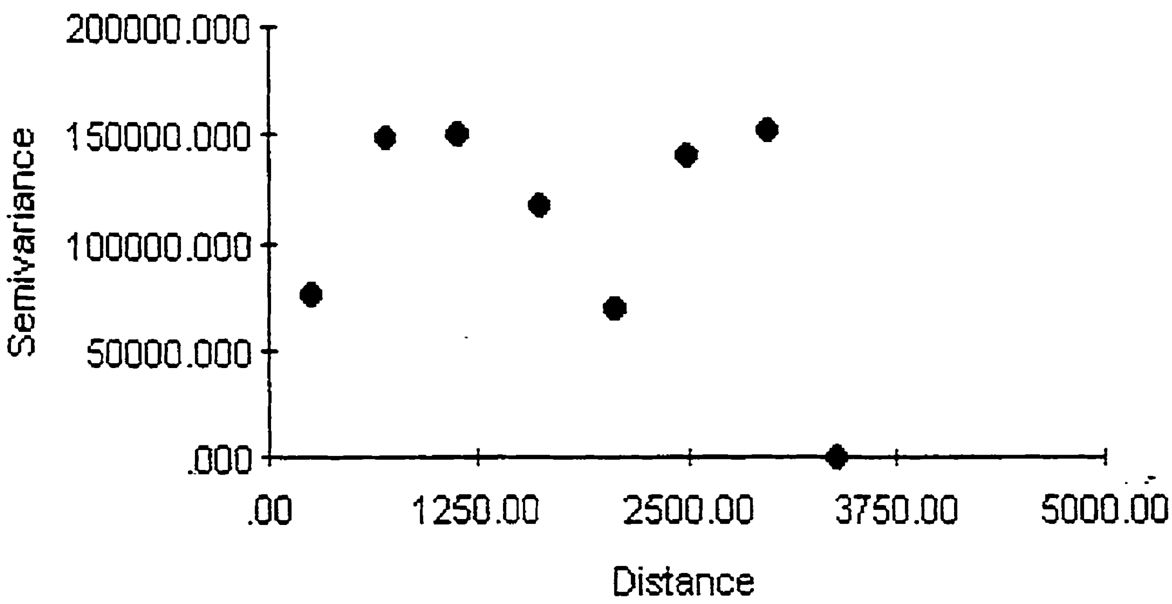


Fig. 17 Variogram for yield Texcoco Watershed, Direction: N-S

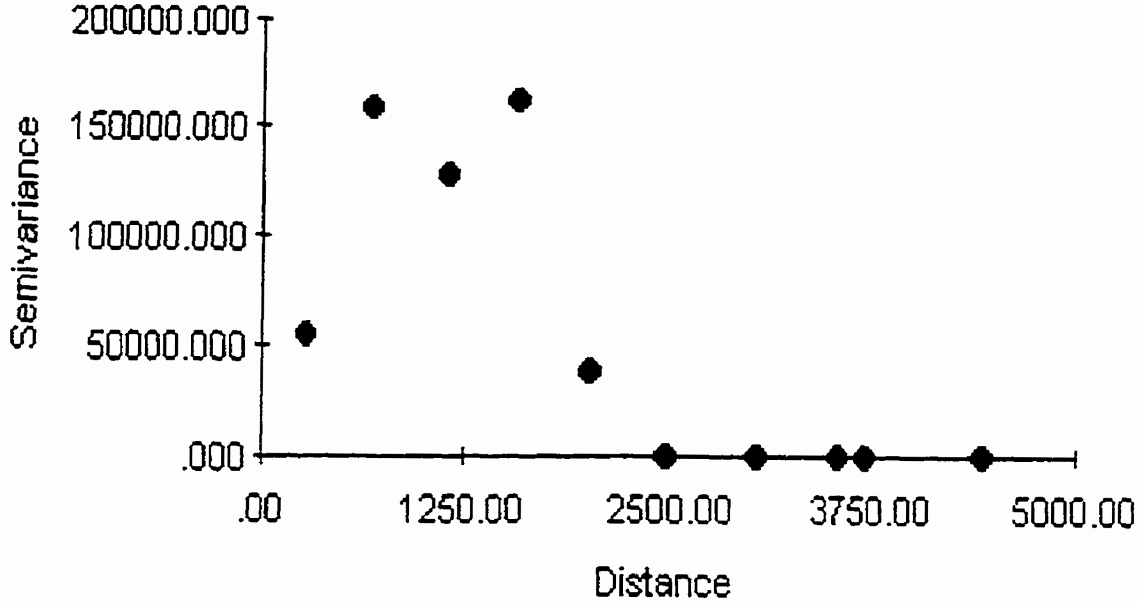


Fig. 18 Variogram for yield Texcoco Watershed. Direction: N-S

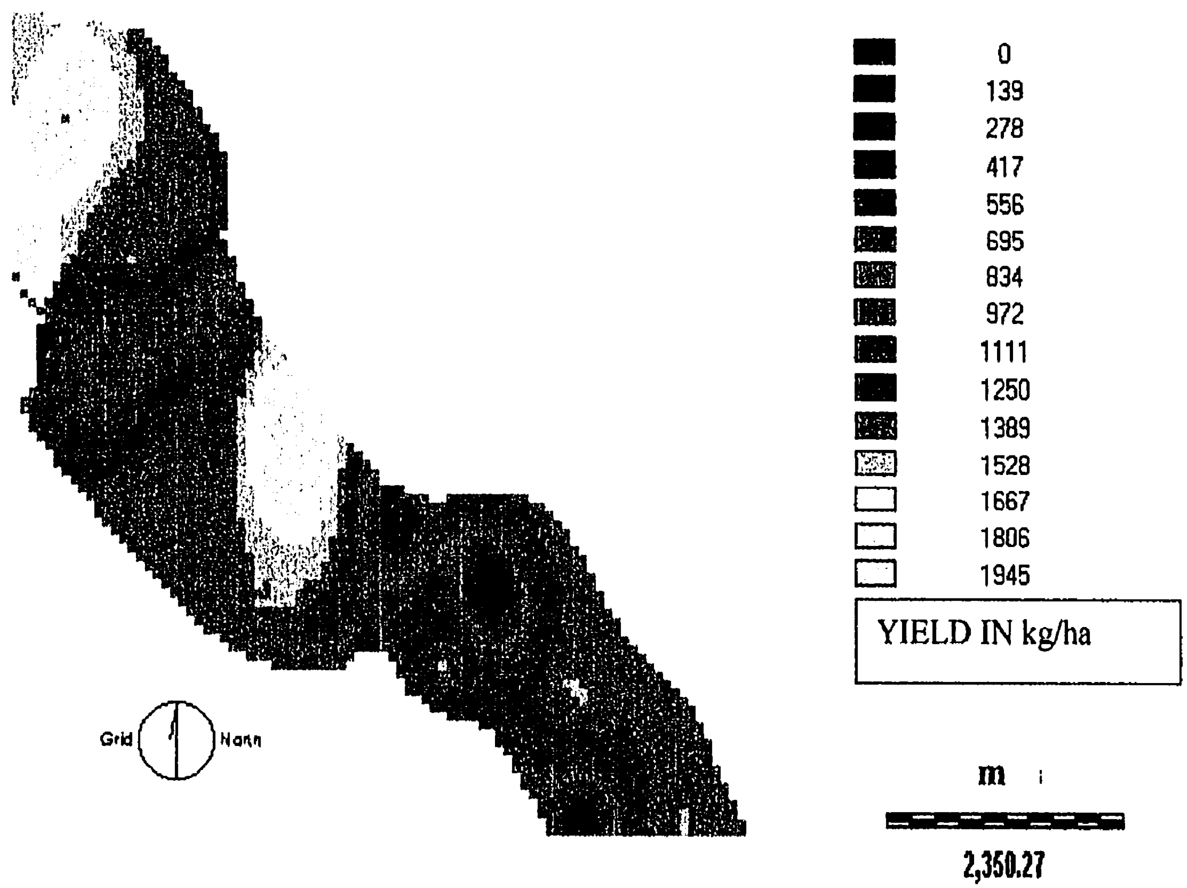
6.5.2 Estimation by Kriging

The values of yield were interpolated by ordinary Kriging at the selected target points. This was done by **"jack-knifing"**, that is, by leaving the target point in turn out of the calculation, estimating it, then moving to the next point while re-introducing the previous point into the calculation until the whole set of target points was estimated. The yield and suitability classes at the target points, once calculated, were used for spatial interpolation by Kriging. Kriging estimates for blocks on a square grid were also computed. The interpolated blocks were set (by default of the software used) to the maximum resolution possible. This turned out to be 78 metres (approximately three quarters of a hectare). The kriged results were written to a grid file. Thus, the blocks forming the grid of interpolated yield values covered the entire study area.

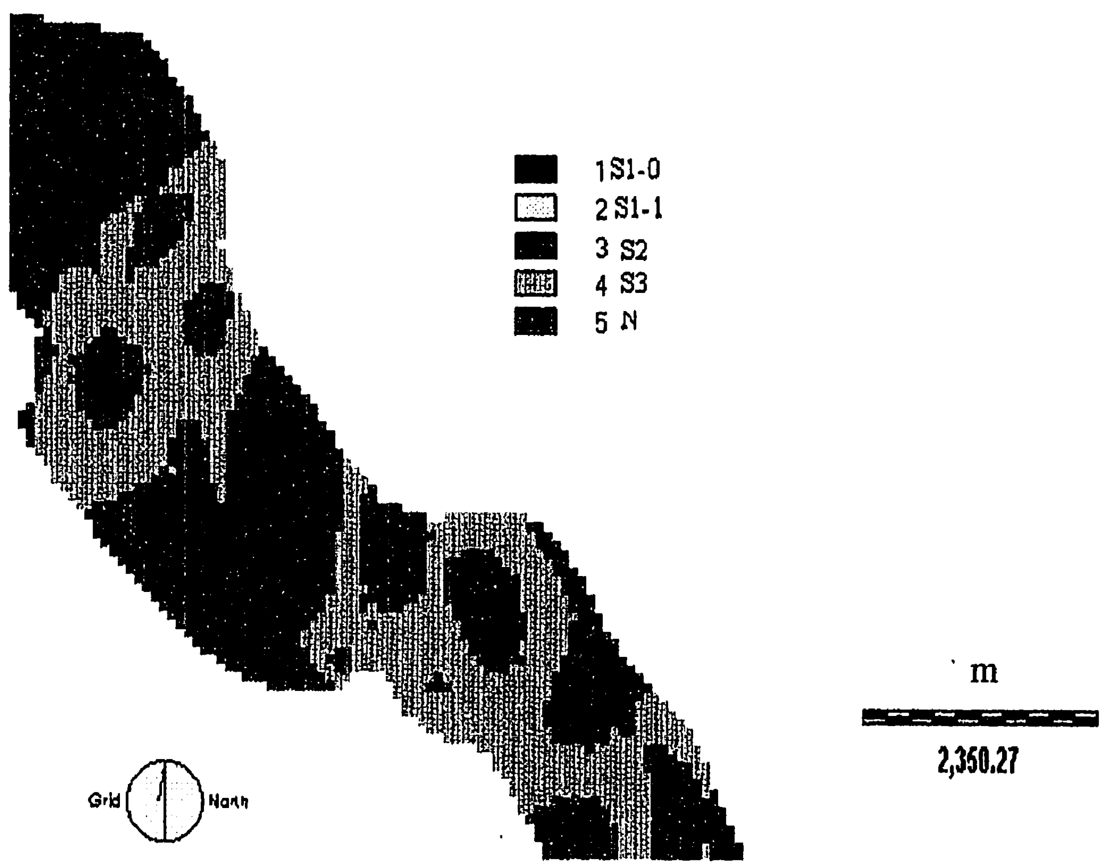
6.5.3 Raster Maps

The grid maps of yield and suitability classes were converted to raster maps using an in-house format-converting program. The resulting image maps were then imported to the ***IDRISI** for Windows* GIS program. A document file containing the number of rows and columns, the reference system and the minimum and maximum **Y** and **X** was created to ensure all the images had the same parameters. The resulting raster map **(Map 7)** shows the predicted yield within the study area. However, to produce a map to illustrate the suitability classes, the raster map of yield was classified using the Reclassify module in ***IDRISI** for Windows* and assigning new values corresponding to the limit of each suitability class **(Map 8)**.

Map 7: Predicted yield by Kriging



Map8: Suitability classes predicted by Kriging



6.6 Fuzzy Mapping

The Fuzzy classification created for yield data was performed in the GIS *IDRISI for Windows*. In this case, 15 Fuzzy membership classes were produced. For this study, the membership function **(Fig. 19)** for each yield class was drawn up to reflect the general yield predicted in the study area. The map produced from Fuzzy analysis **(Map 9)** on the yield data agreed well with that obtained from Kriged analysis **(Map 8).**

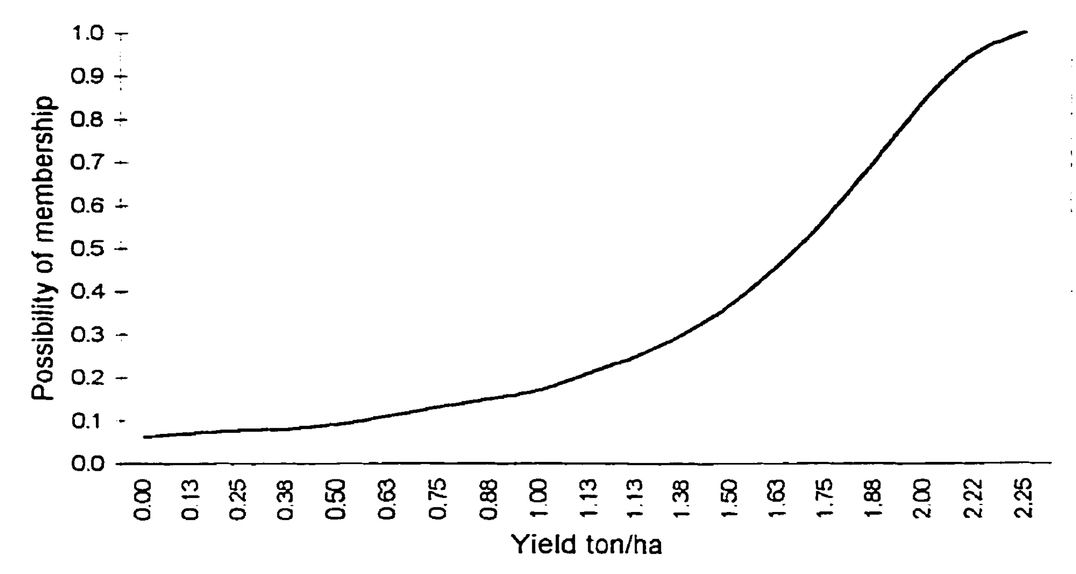
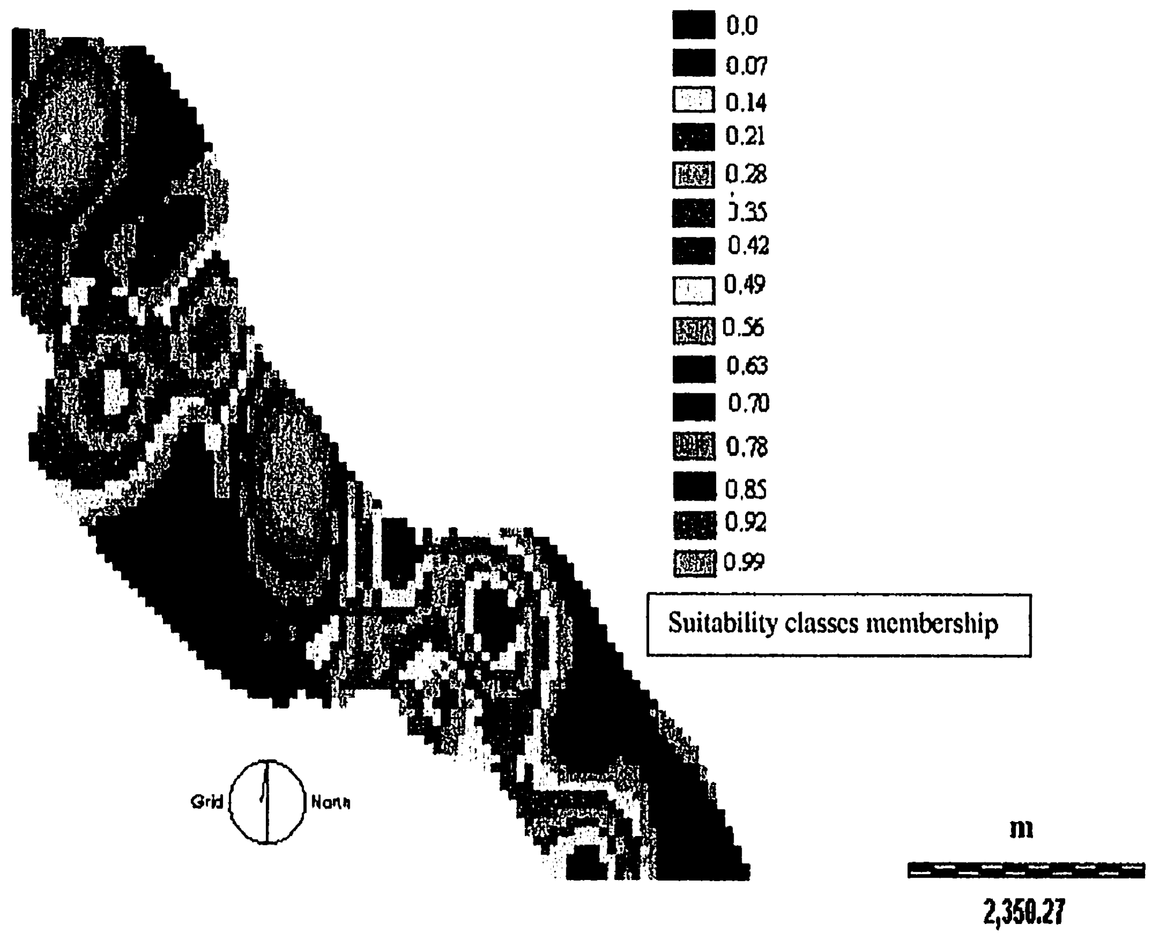


Fig. 19: Fuzzy membership function for suitable land

Map9 : Fuzzy membership classes



6.7 Comparison of Results from Suitability Maps

The estimates of maize (yield) were predicted from suitability classes derived from the three methods: from soil polygons (current paradigm), from spatial interpolation by Kriging and from Fuzzy classification (proposed paradigm). The yield predictions by the three methods were then compared with yield measurements (at 37 target points) and between themselves.

Table 6 shows the estimated values for the three procedures along with their RMS values obtained from comparison with measured values. The results indicate that the Fuzzy map is the most accurate method for prediction. The next most accurate method for land suitability evaluation was the Kriged interpolated values. However, the current method (estimated from polygon map) shows the lowest accuracy compared to the other two methods, with a relatively high RMS value.

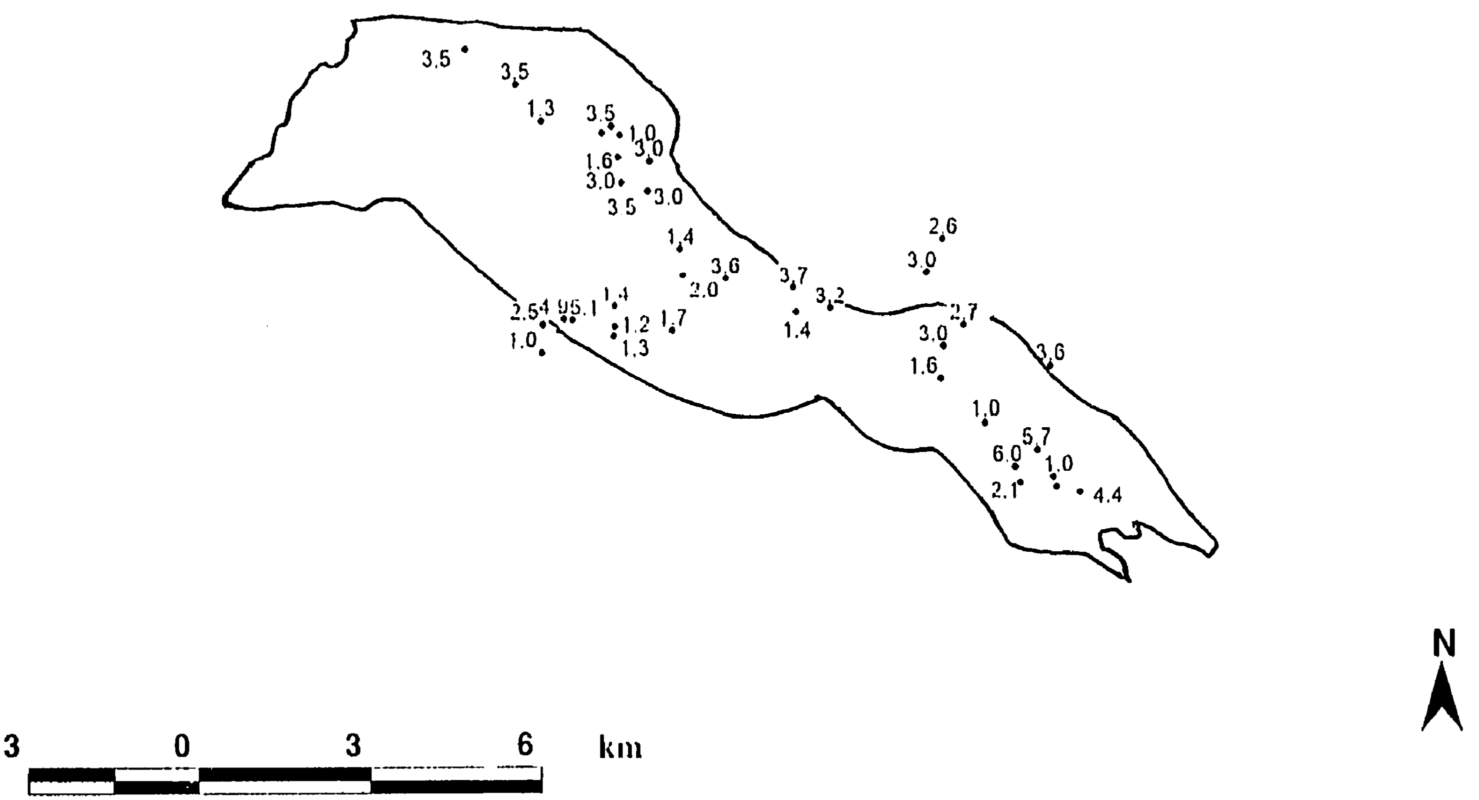
Table 5: Observed and predicted yields at the 37 test sites, Texcoco basin

Easting	Northing	Grain Maize yield by different methods ton/ha			
		Observed	Polygons	Kriged	Fuzzy
517050	2155742	3	0.9	1.74	1.75
521340	2152776	1	0.9	1.76	1.78
521170	2153048	5.7	0.9	1.8	1.84
520935	2152876	6.04	0.9	1.83	1.87
521374	2152672	5.86	0.9	1.78	1.82
520402	2154304	2.7	0.9	2.08	2.16
516704	2154520	1.4	1.8	1.72	1.73
517316	2154259	1.74	0.9	2.08	2.16
516248	2154375	5.1	1.8	1.47	1.27
515911	2154033	1.04	1.8	1.71	1.71
516695	2154315	1.24	1.8	1.84	1.89
516687	2154210	1.34	1.8	1.81	1.86
521324	2153880	3.6	0.9	1.83	1.84
521624	2152612	4.4	0.9	1.77	1.80
516767	2155828	3.5	0.9	1.75	1.76
515911	2156467	1.26	0.9	2.15	2.23
516156	2154387	4.9	1.8	1.42	1.17
515638	2156847	3.5	0.9	2.08	2.16
515925	2154329	2.52	1.8	1.42	1.17
517878	2154824	3.6	0.9	2.26	2.32
515114	2157211	3.5	0.9	2.00	2.08
520000	2154874	3	0.9	1.78	1.80
518618	2154446	1.45	0.9	2.16	2.24
517430	2154855	2.049	0.9	1.79	1.82
517396	2155135	1.44	0.9	1.40	1.13
520174	2155229	2.6	0.9	1.41	1.06
516761	2156316	1.6	0.9	1.8	1.84
516664	2156412	3.5	0.9	1.86	1.91
516734	2156097	3	0.9	1.8	1.84
516566	2156343	1.04	0.9	1.84	1.89
518979	2154493	3.2	0.9	1.8	1.84
518586	2154718	3.74	0.9	2.66	2.13
517069	2156047	3	0.9	1.8	1.84
520617	2153326	1	0.9	1.8	1.84
520183	2154082	3	0.9	1.8	1.87
520158	2153757	1.6	0.9	1.85	1.87
520994	2152722	2.4	0.9	1.8	1.84
		RMS	**117.24**	**38.80**	**38.59**

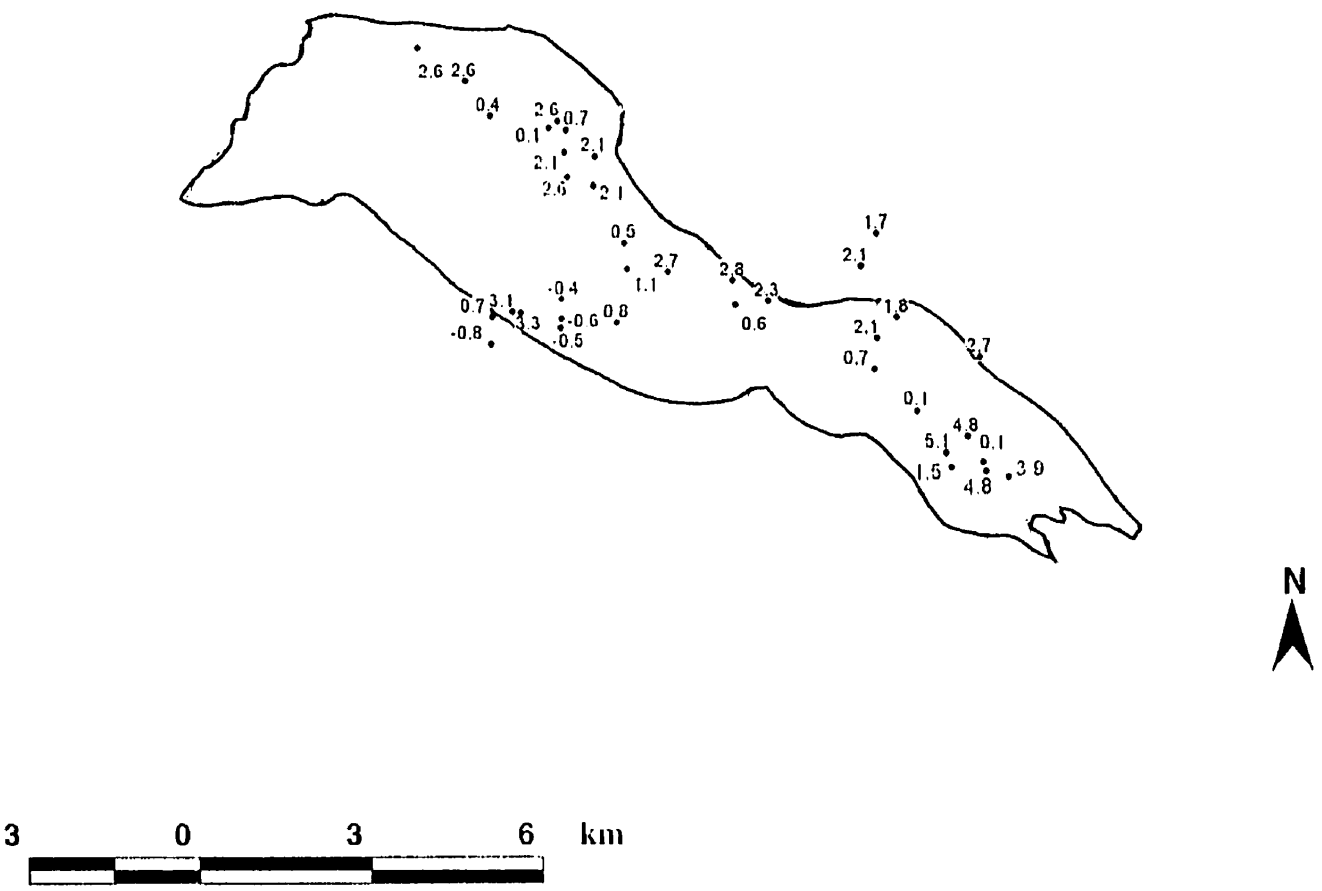
6.8 Spatial Distribution of Deviation of Yield Estimates: Model Calibration

The following maps (**Map 10, 11,12,** and **13**) show the results of plotting the observed yield values and the deviations of the estimates from the observed yields for the two paradigms. These maps show the distribution of the spatial pattern of deviations as indicative of accuracy over the study area. There is no apparent spatial pattern of residuals for any of the predictive methods. **Table 8** shows the results of calculating the percentage of overestimates or underestimates.

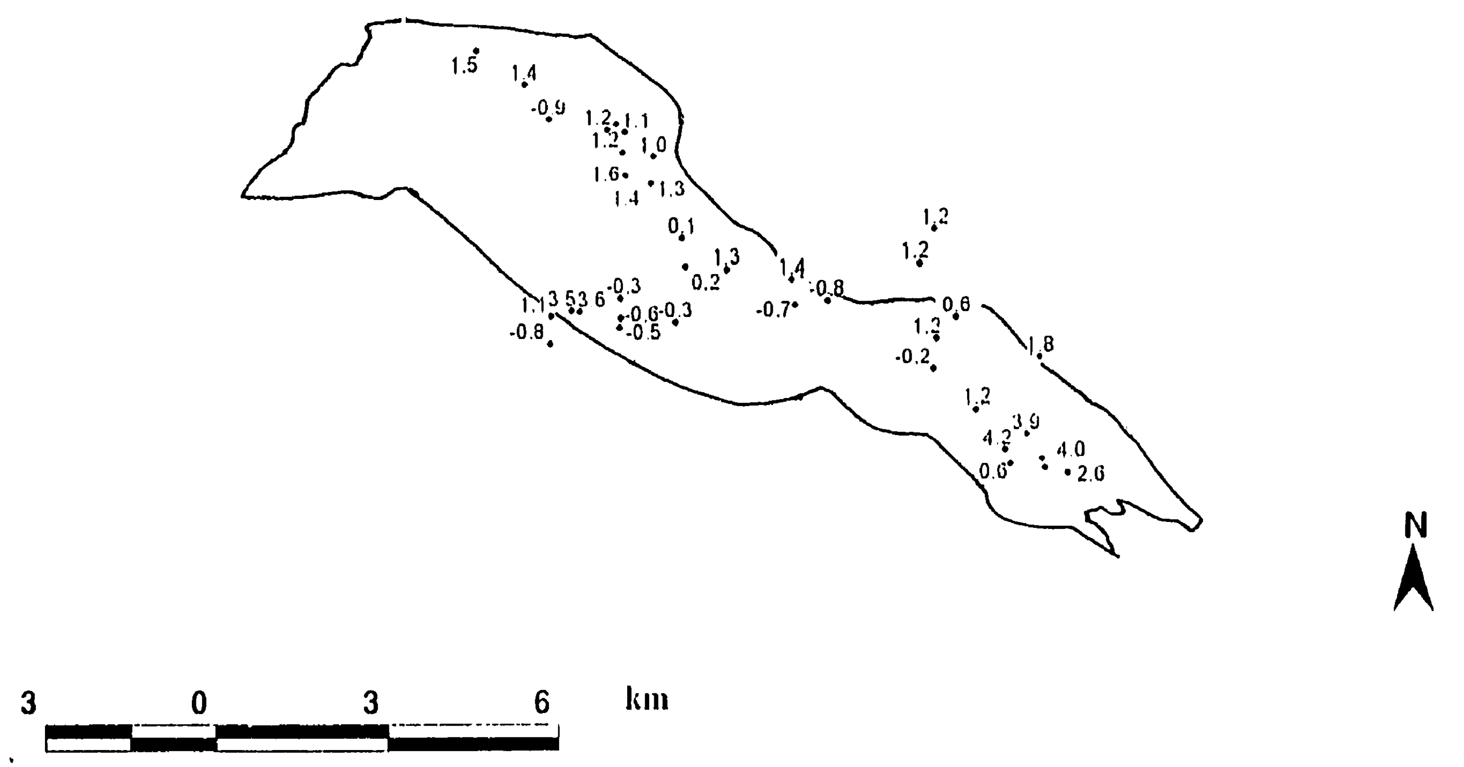
Map 10: Observed (measured) yield data (ton/ha)
at 37 random check sites



Map 11: Spatial distribution of deviations of yield estimates (ton/ha) from the observed. Estimates derived from the evaluation soil polygons (current paradigm)



Map 12: Spatial distribution of deviations of yield estimates (ton/ha) from the observed. Estimates derived from the evaluation at point- data and their spatial interpolation by Kriging (proposed paradigm)



Map 13: Spatial distribution of deviations of yield estimates (ton/ha) from the observed. Estimates derived from the evaluation at point- data and their Fuzzy classification (proposed paradigm)

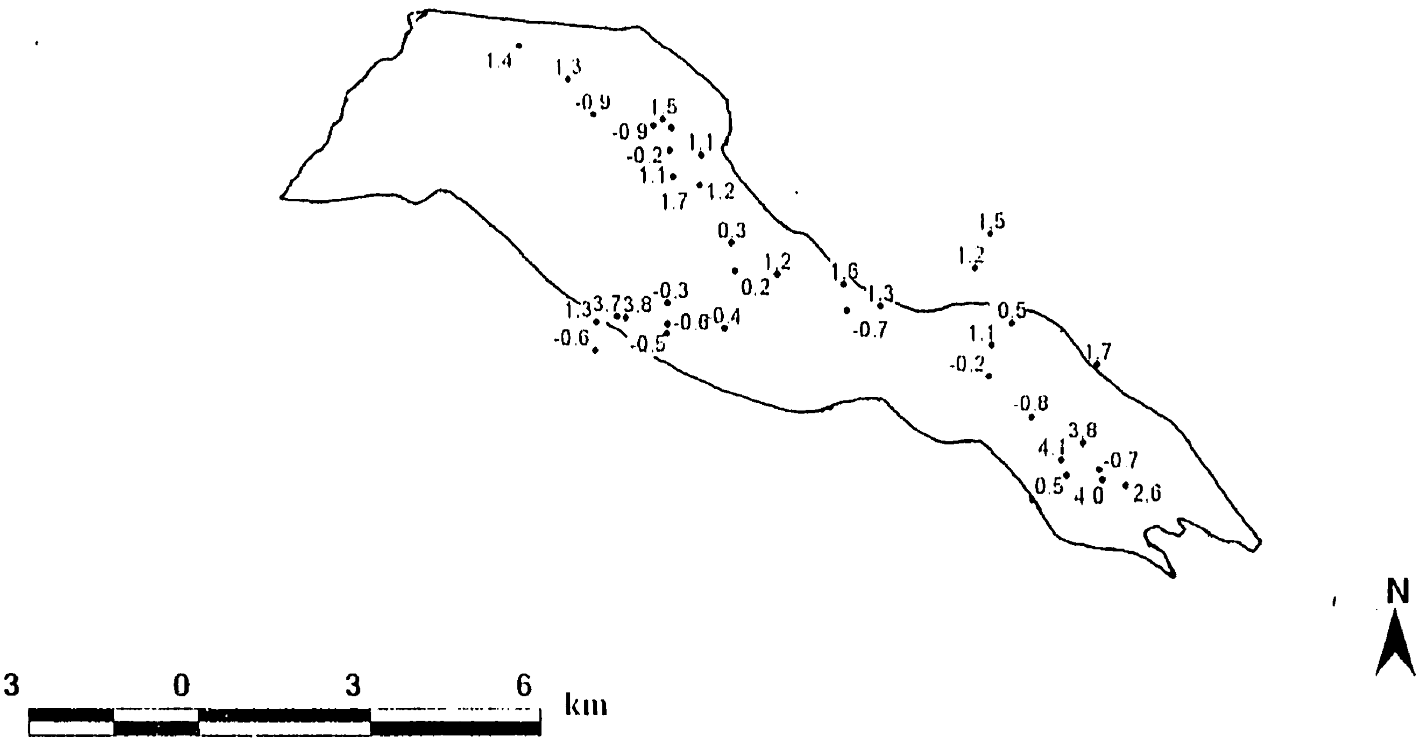


Table 8 The results of the percentage deviations of the three yield predictors from the observed yield

Observed yield (ton/ha)	Soil polygons estimates %	Kriged estimates %	Fuzzy estimates %
1	10	-76	-78
1	70	40	39
1.04	-73	-64	-64
1.04	70	40	39
1.24	-45	-48	-52
1.26	29	-71	-77
1.34	-34	-35	-39
1.4	-29	-23	-24
1.44	56	12	11
1.45	70	41	40
1.6	65	46	59
1.6	70	40	39
1.74	48	-20	-24
2.04	38	-49	-54
2.4	44	-16	-17
2.52	29	44	54
2.6	38	30	22
2.7	67	23	20
3	10	-30	-34
3	70	42	42
3	74	43	41
3	76	29	43
3	74	47	45
3.2	13	-77	-82
3.5	44	-13	-15
3.5	74	41	38
3.5	74	50	50
3.6	75	49	49
3.6	75	37	36
3.74	72	44	43
4.4	80	60	59
4.9	63	71	76
5.1	65	71	75
5.7	84	68	68
5.8	85	70	69
6	85	70	69

7.0 DISCUSSION

7.1 General Discussion

The automated land evaluation model was developed and used as a framework to allow the user to enter any new parameters and data to compute an evaluation for each mapping unit or hard point data.

The land evaluation methodology involves a great number of environmental factors (i.e. soil, climate, landscape, etc.). The variability of these factors in time and space and the approaches used to deal with them significantly affect the results obtained in this research .

An important factor in determining the nature of results obtained is the number of meteorological stations used to derive the climatic parameters to feed the model. Only five had sufficient records to interpolate such parameters spatially. Therefore, the goodness of these interpolations and the algorithms used may be reflected in the accuracy of results, although most of the stations used were very close to the study area.

Moreover, the rating system in the definition of suitability classes is based on knowledge and experience. These are knowledge bases. In this case two sources were used for deriving values for classes: Ponce-Hernandez and Beernart (1991) and Sys (1985). These existing knowledge bases contain information for each land characteristic/ land use requirement and for each LUT. It is not known whether the use of other ranges and threshold values from different knowledge bases would have produced significantly different results. However, it is clear that such results would be consistent in terms of the effect on them by the set of procedures involved in the two predictions tested in this thesis. Furthermore, Hewit and vanWambeke (1982) pointed out that in the application of an existing rating system to other areas, or at different times, it is important to consider the particular land units and the factors on which the transferred system is developed. If regions within national boundaries differ greatly, it may not be possible to develop criteria and limits for an overall national evaluation system. Each region may need a separate treatment.

Another important point is that the spatial variability, represented by polygons in the current paradigm, may not be fairly represented by the selection of the representative profile. When an area is bounded by a line and assigned to a soil mapping unit, the implied assumption according to the selection of the representative profile of such a mapping unit is that the variation over the area of mapping unit is the same or very similar. This may have an effect on the interpretation derived from polygons with internal variability. This is true, since the variation of land resource properties is to be expected within any polygon in the study area. Sometimes, the soil internal properties (e.g., the depth and physical properties) vary even over very short distance within the study area. In this case, for land evaluation purposes, it is necessary to produce valid generalizations about areas relevant to particular land use issues.

7.2 Discussion on Regression Analysis

The exploratory regression and correlation analysis gave only marginally encouraging results. This analysis was undertaken in order to estimate missing climatic data and so to be able to increase the accuracy of predictions using these data by the evaluation models within the study area. From these results introduced in section 6.1 a high correlation was found between annual rainfall and rainfall growing season, with a coefficient determination of 0.98. Other relationships explored between annual rainfall and for instance, elevation did not yield any significant correlation with coefficient of 0.11. In order to predict the data for length of growing period, the analysis performed found that there is no significant correlation of this variable with none of annual rainfall, rainfall within the growing season and with elevation. The coefficients are 0.08, 0.09 and 0.2 respectively. Thus, only one regression equation (annual rainfall with rainfall growing season) was used in the estimation of missing climatic data. The number of "usable" meteorological stations for annual rainfall became 12, (five with original data and 7 with estimated data by the regression equation). These enhanced the results of spatial interpretation of climatic parameters for the suitability assessment model.

7.3 Discussion on Spatial Analysis and Interpolation

Analysis of spatial dependence of suitability classes represented by predicted yield for the study area, using semivariograms, indicated that variation of suitability classes and predicted yield was generally isotropic. Directional semivariograms in the **North-South** and **N-east-S-west** directions indicated that the variation become irregular at a distance of **2000 m** and **4000 m** due to the shape and size of the study area. The dominant limiting climatic factors were temperature and rainfall. These two variables change from east to west. Temperature increases with decreasing altitude and rainfall decreases to the west with decreasing altitude in the middle of the watershed. The isotropic semivariograms were estimated directly from the data and the Gaussian semivariogram model turned out to be the most accurate. On the other hand, cross-validation results for the selected semivariogram did not appear to show significant differences between the RMS values for the two selected models for the omnidirectional semivariogram. This indicates that there are no preferential directions of regionalization of crop yields in the study area.

Block kriged values of suitability ratings and yields in **Map 7** ranged from 550 to 1945 kg/ha for grain yield values. This wide range of values is evidence of the intrinsic variability of soil and climatic factors that determine crop yields over the study area. In most of the points that were evaluated, the classes tended to be **S3** and **S2** for the limiting factors other than climatic factors. However, some points were controlled by salinity and slope, and while others were limited by fertility and other factors.

It must be noted that the results may be strongly influenced by the characteristics and intrinsic limitations of the suitability rating knowledge bases tapped into. For instance in some cases the mean minimum temperature was **9^0 C.** This value falls exactly on the border between class **S2** (with range between **7-9 ^{0} C**) and class **S3** (with range between **9-12 ^{0} C**). However, based on a **9^0 C** value the land would be classified as **S3,** which seems not to make much intuitive sense. On the other hand if there is a temperature reading of **8.5 ^{0} C** then this would be classified as **S2** even though this is only **0.5** degree away from the higher class. So, the "crispness" of classification may be a hindrance to accuracy of interpretive results.

It most be noted that the differences between estimates produced from applying any of the interpolation methods for climatic data do not have any effect on the results obtained from the suitability assessment models in ALES. This is because the decision trees constructed in the model use as nodes of the tree sufficiently broad ranges for each suitability class so as to allow for differences in interpolated climatic values without a change in the resulting suitability class after applying the decision tree models.

7.4 Discussion on the Soil Information

In this section the two paradigms will be discussed and the comparison between the three different methods of suitability evaluation will be addressed.

7.4.1 Land Evaluation from Hard Point Data

Conventional methods of providing soil information (current paradigm) are based on the published soil map, and the information that can be retrieved from them is at a higher level of aggregation, the level of map units. Predictions at points can be derived from a soil map, but they are equal for all points in the same map unit (Burrough, 1998; De Gruijter et al., 1997). Therefore, at best the prediction consists of averages of any soil property over the entire area of the mapping unit

The methodology of land evaluation used in this study as a part of the proposed paradigm is needed to produce a suitability map. This consisted of the combined use of the Regionalized Variables Theory and Fuzzy set theory to enable mapping of the study area as a continuous surface.

The evaluation, interpolation and continuous classification strategy that has been followed aimed at evaluating all point observations and predicting the total suitability for the study area, in order to avoid any data generalization when changing from traditional crisp or discontinuous classification (current paradigm) to the continuous one. In other words: "evaluate first, interpolate later" instead of " interpolate first and evaluate later". It was possible in this study to ascertain what would be the effect on results if the order of procedures would have been inverted. A disadvantage of this strategy may be that the information about the spatial distribution for each variable involved in the assessment is being included within the others in the rating. The opposite would be to interpolate every variable independently, then create a thematic raster map for each variable (pH, O.M, EC, etc.) then overlay them on top of each other to create the suitability map by a combination of such attributes. This option was not explored due to its computationally intensive nature, as well as time limitations.

In spite of the difficulties faced by the procedures in the proposed or "new" paradigm of this thesis, it became evident that this approach is very useful since the actual hard point data are retained, interpolated and Fuzzified. This is so, even after considering that only the results of evaluating suitability were interpolated and not the actual variables needed to carry out the interpolation.

7.4.2 Land Evaluation from Generalized (Polygon) Data

The philosophy behind the suitability classification (FAO, 1976) is that the polygons defined for the study area are homogeneous over the mapping unit. The polygons were evaluated by extracting the soil characteristics that are necessary as input into the decision-tree models, from the legend of the soil map. Typically, the legend reports data in terms of a soil class. The soil class has a central concept, the typical profile, that characteri zes the whole area covered by the polygon. So, a given suitability class derived from the assessment of the typical profile and class, applies to the whole extent of the polygon with no regard for the possible internal variability of such polygon. Further, if adjacent polygons end up with the same assigned suitability class, they will all into one larger area with the same suitability class.

The most important issue in changing paradigms was to avoid the loss of information during the two processes of generalization mapping and classification. It must be mentioned too that, due to the scope and time limitations of this research, interpolating the suitability class resulting from evaluating point-data creates the problem of having to assign yield intervals (classes) to such assessments in order to convert them to a ratio scale. The errors accrued by this procedure might have decreased accuracy. However, in spite of such errors, the procedures in the proposed paradigm were superior to these obtained using generalized data in the form of mapping units for land suitability assessment.

7.4.3 Comparison

The Kriged and Fuzzy maps **(Map7, 8 and 9)** are completely different from the raster map derived from rasterizing the polygons part of the current paradigm **(Map 6)** if they are compared in terms of dissimilarity of patterns. On the one hand, the information used to produce the polygon map was taken from a soil map published in 1978. The representative soil profile was used to estimate the values of the soil characteristics within each polygon. On the other hand, the other two maps were produced by retaining the hard point data from the field. However, it should be noted that the maps were produced in two different GIS programs, so the results are dependent on the algorithms used by such programs.

The Kriged and Fuzzy maps have a greater resemblance to one another than to the polygon map. Only few areas in the Fuzzy map appear to have an extension in some areas between the classes boundary. especially in the south east of the study area. However, the Fuzzy suitability classes **(Map 9)** and the polygon **(Map 6)** show a different representation of the area. Where the polygon map has only two classes. the Fuzzy map has more classes and shows a greater variation of such classes in the study area.

In order to study the quality of the maps and their efficacy as mechanisms for land evaluation, the yield results were compared with observed grain maize yield. The **RMS** values obtained from the polygon map, Kriged map and Fuzzy maps were **117.24**, **38.80** and **38.59**, respectively. The spatial variability of residuals from the observed yield values by the three techniques used as predictors (i.e. soil polygon, Kriging and Fuzzy classes) allowed for the elucidation of the spatial pattern of predictive behaviour by the models over the studied area. The spatial distribution of yield residual from the observed did not appear to have a pattern. A test of randomness of such residuals falls beyond the scope of this thesis. However, it can be noted that the soil polygon map (current paradigm) tends to under estimate observed yields (see **Table 8**). These underestimates are comparatively high as related to the two other predictive methods. The residuals from estimates by Kriging and Fuzzy classes show a slight tendency to overestimate small yields and underestimate high yields. However, a barely noticeable (and perhaps may be significant) pattern can be noted on the three maps. The cluster of point-data in the south-central portion of the area studied has a slight tendency to yield overestimates by all three predictive methods.

In the light of the results obtained in this study, and in spite of its shortcomings, there is substantial evidence that the components of the proposed paradigm, i.e. Fuzzy set and kriging interpolation methods. are superior for implementing suitability evaluations than the exact polygon approach. However. both Fuzzy set theory and interpolation methods produced similar results over the study area.

8. CONCLUSION

The major aim of this research was to introduce and apply a relatively new paradigm to soil information and land suitability assessment, which does away with the need for (generalized) soil information in the form of soil classes and mapping units as represented by polygons. This new paradigm consisted of retaining non-generalized information and applying geostatistical spatial interpolation and fuzzy boundary representation through membership functions, in order to develop a final map of land suitability classes for maize in the Texcoco watershed of central Mexico.

In light of the evidence found in this study, it can be concluded that the first hypothesis formulated and introduced by this research is rejected. Hence, there are significant differences in the accuracy of estimates of land performance, as predicted by the techniques part of the new paradigm proposed in this study, when compared with estimates derived from the conventional paradigm consisting of generalized information- i.e. crisp soil classes and soil mapping units (polygons). Suitability classes derived from retaining "hard" point-data, interpolation and Fuzzy boundaries represent a significant improvement in accuracy.

When comparing the accuracy of performance estimates (yield) from the application of Fuzzy membership functions to classes against those obtained from the application of Kriging interpolation alone, no significant differences in accuracy, as indicated by the Residual Mean Square of predictions, were found. Hence, the second hypothesis formulated and introduced by this research is not rejected.

It can be concluded that there are considerable advantages, in terms of accuracy of interpretations derived from soil data, by retaining hard point data in soil databases and then using a suite of algorithms for spatial interpolation and for Fuzzy membership and boundary representation to derive interpretive maps. Geostatistical techniques and Fuzzy Set Theory and algorithms, used in a Geographical Information System (GIS) environment resident in modern computer technology, are now sufficiently powerful tools to prevent and to avoid unnecessary soil information losses due to generalization.

As proven in some instances (Ponce-Hernandez, 1994), an alternative approach may be to produce thematic raster maps for each variable (e.g. salinity, drainage, or any soil elements) generated by the application of interpolation to raw-point-data. These maps can be then analysed further using modern GIS software allowing for spatial modelling and derivation of informative results

BIBLIOGRAPHY

Aronoff, S. 1995. Geographic Information Systems: A Management Perspective. WDL Publications. Ottawa, Canada. 294pp.

Askew, H.O. and Rigg, T. 1932. The importance of depth of soil sampling in studies relating to the mineral content of pastures. N.Z.Jl. Sci. Technol. 13(284)

Atkinson, P.M. Curran, P.J. and Webster, R. 1990. Sampling remotely sensed data imagery for storage, retrival and reconstruction. Professional Geographer . 42: 345-353

Avery, B.W. 1987. Soil Survey Methods: A Review. Technical Monograph no.18. Silsoe: Soil Survey & Land Resource Centre. 86 pp.

Bascomb, C.L. and Jarvis, M.G. 1976. Variability in three areas of the Denchworth soil map unit J. Soil Sci. 27: 420-437.

Becket, P.H.T. 1984. Soil Information Systems: the problem of developing a capacity of reappraising data for single purposes. In Soil Information Systems. A Proceeding of the Sixth Meeting of the ISSS Working Group on Soil Information Systems. Bolkesø, Norway. 28 Feb-4 Mar. 1983.Pudoc Wageningen

Becket. P.H.T. and Webster. R. 1971. Soil variability: A review. Soils and Fertilizers. 34:1-15

Black, C.A. 1965. Methods of Soil Analysis. Part 2. Agronomy Monograph No.9, ASA, Madison, Wisconsin. U.S.A.

Bouma, J. 1989. Land qualities in space and time. In Land qualities in space and time Ed. by Bouma, J. and Bregt, A.K. Pudoc Wageningen. 352 pp.

Buol, S.W. Hole, F.D. and McCracken, R.J. 1989. Soil genesis and classification. 3rd.Ed. Iowa state University. Press, Ames

Burgess, T.M. and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties I. The semi-variogram and punctual Kriging. J. Soil Sci. 31: 315

Burgess, T.M. and Webster, R. 1984. Optimal sampling strategies for mapping soil types. I.Distribution of boundary spacings. J. Soil Sci. 35:641-54.

Burnham, C.P. 1986. Classification and maps of soils and their interpretation for Planners. In Land and Its Uses-Actual and Potential: An environmental appraisal. Ed.By Last, F.T. Hotz, M.C.B. and Bell, B.G. Plenum Press. New York and London. 597pp.

Burrough, P.A. 1983. Multiscale sources of spatial variation in soil.
I . The application of fractals concepts to nested levels of soil variation
J. of Soil Sci. 34: 577-597

Burrough, P.A. 1987. Mapping and map analysis: new tools for land evaluation.
Soil Use and Management 3: 20-25.

Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation
J. Soil Sci. 40:477-492.

Burrough, P.A. 1992. The development of intelligent Geographical Information System.
Int. J. Geogr. Inf. Systems.6:1-12.

Burrough, P.A. Beckett, P.H.T. and Javis, M.G. 1971. The relations between cost and
utility in soil survey. II. Conventional or free survey. J. Soil Sci. 22: 369-381.

Burrough, P.A. Brown, L. and Moris, E.C. 1977. Variation in vegetation and soil pattern
across the Hawkesbury Sandstone Plateau from Barren Grounds to Fitzory Falls.
N.S.W. Australian J. Ecol. 2: 187-159.

Burrough, P. A and. Frank, A.U. 1996. Geographic objects with indeterminate
boundaries.Taylor & Francis. London 345p.

Burrough, P.A. and McDonnell. R.A. 1998. Principles of Geographical Information Systems. Oxford University Press. New York .333 p.

Burrough, P.A. MacMillan, R.A. and van Deursen, W. 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. J. Soil Sci.43:193-210

Butler, B.E. 1980. Soil Classification for Soil Survey. Clarendon Press, Oxford.

Chang, L. and Burrough, P.A. 1987. Fuzzy reasoning: a new quantative aid for land evaluation. Soil Survey and Land Evaluation 7: 69-80.

Cline, M.G. 1949. Basic principles of soil classification. Soil Sci. 67: 81-91

Cook, S.E. Corner, R.J. Grealish, P.E. and Charters, C.J. 1996. A rule-based system to map soil properties. Soil Sci. Am. J. 60:1893-1900

Dahiya, I.S. Richter. J. and Malik, R.S. 1984. Soil spatial variability; Review. Inter. J. Trop. Agri. II: 1-102

David, L.R. 1994. Soil Science: methods and applications. Longman. 350 pp.

David, M. 1977. Gesostatistical Ore Reserve Estimation. Elsevier, Amsterdam. 364p.

Davidson. D A. 1992. The Evaluation of Land Resources. Longman Scientific and Technical. 198p.

Davidson, D.A. Theochropoulos, S. P. and Bloksma, R.J. 1994. A land evaluation project in Greece using GIS and based on Boolean and fuzzy set methodologies. Int. J. Geographical Information Systems. 8(4): 369- 384

De Gruijter, J.J. and McBratney, A.B. 1988. A modified Fuzzy *K*-means method for predictive classification. In: Bock, H.H. (ed.), Classification and Related Methods of Data Analysis. Elsevier, Amsterdam. 97-104.

De Gruijter, J.J. Waloort, D.J. and van Gaans, P.F.M. 1997. Continuous soil maps- a Fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77: 169-195

Dent, D. and Young, A. 1981. Soil Survey and Land Evaluation. George Allen & Unwin London, England.

Direccion General De Geografica. A983. Soil map 1:50 000. SPP. Programacion y Presupuesto.

Direccion General De Geografica. A983. Orthophoto map 1:20 000. SPP. Programacion y Presupuesto.

Eastman, J.R. 1997. "Idrisi for Windows: User's Guide Version 2.0". Worcester. Clark University.

Englund, E. and Sparks, A. 1991. " Geostatistical Environmental Assessment of Software:Users guide". Nevada: U.S. Environmental Protection Agency.

Eyk, J.J. Macvicar, C.N. and Villiers, J.M. 1969. Soils of the Tugela Basin: a study in subtropical Africa. Natal Town & Regional Planning Reports 15. Natal Town & Regional Planning Comission. 263

FAO. 1974. FAO-UNESCO Soil Map of the World 1: 5000 000. Unesco-Paris. 55p.

FAO. 1976. A Frame Work For Land Evaluation. FAO Soils Bulletin No. 32. ILRI Publication No.22. Italy: Food and Agriculture Organization of the United Nations. 87p.

FAO. 1978. Report on the Agro-ecological zones project. World Soil Resources Report 48, Vol. 1: Methodology and results for Africa. Rome: FAO. Xi, 158 p.

Gerasimov, I.P. 1983. Land resources of the world; their use and reserves. Definitions of land quality and agriculture potential by soil survey maps. Geoforum 14:427-439.

Giltrap, D.G. 1981. A practical application of spatial interpolation for the rapid production of a wide range of special purpose soil maps. Information Systems for soil and related data. Proceedings of the 2nd Australian meeting of ISSS Working Group on Soil Information Systems. Canberra, Australia, 73-82

Giltrap, D.J. 1983. Computer production of soil maps I. Production of grid maps by Interpolation. Geoderma. 29: 295-311.

Giltrap, D.G. 1983. MIDGE a microcomputer Soil Information System. In "Proceedings of the 6th Meeting of ISSS Working Group on Soil Information Systems, Bolkesj, Norway.112-119

Goovaerts, P. and Journel, A.G. 1995. Integrating soil map information in modelling the spatial variation of continuous soil properties. European Journal of Soil Science 46: 397- 414

Hendricks Franssen, H.J.W.M. van Eijnsbergen, A.C. and Stein, A. 1997. Use of spatial classification for mapping soil pollutants. Geoderma 77: 243-262

Hewitt, A.E. and Van Wambeke, A. 1982. Soil, land and site evaluation methods for Land use planning . First Int. Symposium on Soil, Geology and Landforms Impacton Land Use Planning in Developing Countries. Bangkok, Thiland 1-3, April1982.

Hole. D. F. and Campbell. J.B. 1985. Soil Landscape Analysis. Rowman and Allanheld. Totowa, NJ. 196 pp.

Huajun, T. ,Debaveye J., and Da Ruan, Ranst E. Van . 1991.Land suitability classification based on fuzzy set theory. Pedologie,XLI:227-290.

ILWIS 2.1 for Windows. 1997. "The Integrated Land and Water System: User's Guide. ILWIS Department. International Institute for Aerospace Survey & Earth Sciences. Enschede. The Netherlands.

INEGI. 1986. Cartas edafologica, geologica y uso potencial del suelo. Escala 1:250000 del Estado de Mexico.

Isaaks, E.H. and Srivastava, R.M. 1989. Applied Geostatistics. Oxford University Press. New York, Oxford. Oxford University Press, Inc. 561pp.

Journel, A.G. 1996. Modelling uncertainty and spatial dependence: Stochastic imaging. Int. J. Geographical Information Systems. 10:517-522.

Journel, A.G. and Huijbregts, C.J. 1978. Mining Geostatistics (London: Acadmic Press).

Karla, Y.P. and Maynard, D.G. 1991. Methods Manual for Forest Soil and Plant Analysis. Information report. NOR-X-319 Forestry. Canada, Northwest Region, Edmonton, Alberta.

Kanty, B.A. and Morse, R.K. 1965. A modern approach to highway material sampling. Proc. 6th In. Conf. Soil Mech. Fdn. Engng. 1/12, 55-58.

Lagacherie, P. Cazemier, D.R. van Gaans, P.F.M. and Burrough, P.A. 1997. Fuzzy *k*-means clustering of fields in an elementary catchment and extrapolation to a larger area. Geoderma 77: 197-216

Lathrop, Richard G. Jr., Aber, John D., and Bognar, John A. 1995. Spatial variability of digital soil maps and its impact on regional ecosystem modeling. Ecological Modelling. 82:1-10.

London, J.R. 1991. Booker Tropical Soil Manual. Booker Tate. Thame, England. 474pp.

Lyford, W.H. 1974. Narrow soils and intricate soil pattern in South New England Geoderma 11: 195-208

Marquez-Rodiles, F. 1990. Zonificacion Agroecologica De Maiz (zea mays) De Temporal En El Estado De Mexico. Thesis Profesional. Departmento De Suelos. Universidad Autonoma Chapingo.

Matheron. G. 1971. The theory of regionalized variables and its applications. Cahiers du Centre de Morphologie mathematique de Fontainebleau. No.5. Paris 212 p.

Mazaheri, S.A. Koppi, A.J. and Mc Bratney, A.B. 1995. A fuzzy allocation scheme for the Australian Great Soil Groups Classification system. European J. of Soil Sci. 46:601-612

McBratney, A.B. and DeGruijter, J.J. 1992. A continuum approach to soil classification by modified fuzzy *K*-means with extragrades. J. Soil Sci. 43:159-175.

McBratney, A. B. De Gruijter, J. J. and Brus, D. J. 1992. Spatial prediction and mapping of continuous soil classes. Geoderma. 54:39-64.

McBratney, A. B. De Gruijter J. J. Brus, D. J. 1992. Spatial prediction and mapping of continuous soil classes. Geoderma 54: 39-64

McBratney, A. B. and Odeh, I. O.A. 1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77: 85-113

McBratney, A. B. and Webster, R. 1983. How many observations are needed for regional estimation of soil properties. Soil Sci. 135: 177-183

McBratney, A. B. and Webster, R. 1986. Choosing function for semi-variogram of soil properties and fitting them to sampling estimates. J. Soil Sci. 37:617-39.

McKeague, J.A. 1978. Manual of Soil Sampling and Methods of Analysis. Canadian Society of Soil Science. 212pp.

Northcote, K.H. 1979. ' A factual key for the recognition of Austrailian Soils.' CSIRO Division of Soils Divisional Report 2/65.

Odeh, I. O. A., McBratney A. B. and Chittleborough D. J. 1992. Fuzzy-*C*-means and Kriging for mapping soil as a continuous system. Soil Sci. Soc. Am.J. 56:1848-1854.

Odeh, I. O. A., McBratney A. B. and Chittleborough D. J. 1992. Soil pattern recognition with Fuzzy-*c*- means: application to classification and soil-landform interrelationships. Soil Sci. Am. J. 56: 505-516

Oliver, M.A. 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geographical Information Systems. 4:313-332.

Parson. J.R. 1971. Prehistoric Settlement Patterns in the Texcoco Region. Mexico. Memoris of the museum of Anthropology, University of Michigan. no. 3. Ann .Arbor

Ponce-Hernandez, R. 1994. Improving the representation of soil spatial variability in geographical information systems: A paradigm shift and its implications. p. 780-800 in 15th World Congress of Soil Science. Transactions. Volume 6a, Commission V: Symposia.872pp.

Ponce-Hernandez, R. 1998. Ecological and Economic Zoning: A proposed Methodology For the Amazon Basin. Project FAO GCP/RIA/128/NET.

Ponce-Hernandez, R. 1995. Land Resources Inventories for Land Evaluation and Land Use Planning: a Combined Approach for Mexico. p. 65-88 in Land Evaluation for Sustainable Agriculture for Mexico. By Ruiz-Figueroa., J.F. Primera edicion.

Ponce-Hernandez, R. and Beernart, F. 1991. Manual of Land Evaluation for Rainfed Agriculture. I.N.I.A./FAO/ UNDP/86010. Maputo, Mozambique.

Ponce-Hernandez, R. and Biosse, A.J. 1991. Automated Land Evaluation in Mozambique:The Machava Polot Area, Maputo Province; Systems Trials and Results.FAO/UNDP MOZ/86/010. Maputo. Mozambique.

Raupach, M. 1951. Studies in the variation of soil reaction. Aust. J. Agric. Res. 2: 83-91

Ragg, J.M. and Henderson, R. 1980. A reappraisal of soil mapping in an area of southern Scotland. J. Soil Sci. 31:559.

Rivera-Vargas, M. 1993. Determinacion De La Aptitud Fisca De Maiz(Zea Mays) Y Frijol (Phaseolus Vulgaris) Para El Municipio De Tezecoco. Chapingo, Mexico.

Robinson, G.W. and Lloyd, W.E. 1915. On the probable error of sampling in soil survey. J. Agric. Sci. Camb. 7: 144-153.

Rossiter, D.G. 1989. ALES: A Microcomputer program to assist in land evaluation In Land qualities in space and time. Edited By Bouma, J. and Bregt, A.K. Pudoc Wageningen.p 113-115.

Rossiter, D.G. 1998. Lecture Notes: Methodology for Soil Resource Inventories. SOL.27. International Institute for Aerospace Survey& Earth Sciences(ITC).

Rossiter, D.G. 1994. A Compendium of On-Line Soil Survey Information: Soil Classification. Internet: http://www.itc.nl/~rossiter.

Rossiter. D.G. and Van Wambeke. A.R. 1995. "Automated Land Evaluation System: ALES Version 4.5 User's Manual". Cornell University.

Sandres, W.T. Jeffery, R.P. and Stanley, R.S. 1979. The basin of Mexico: Ecological Process in the Evolution of Civilization. Academic Press; N.Y.

Soil Survey Staff. 1976. Soil Taxonomy. US Government Printing Office. Washington, D.C.

Sys, C. 1985. Land evaluation. Ghent, Belgium: State University of Ghent, Int. Training Centre for post-graduate soil scientists. Algemen Bestuur van de Ontwikkelingssamenwerking.

Sys, C., van Ranst, E. and Debaveye, J. 1991. Land evaluation, part II: methods in land Evaluation. ITC, University of Ghent, Belgium. 247 P.

Tabor, J.A., A.W. Warrick, D.A. Pennington, and D.E. Myers.1984. Spatial variability of nitrate in irrigated cotton: I. Petioles. Soil Sci. Am. J. 48: 602-607.

Tang, H., Debaveye, J. Da, R. and van Ranst, E. 1991. Land classification based on fuzzy set theory. Pedologie, XLI:277-290.

Tang, H.J., and van Ranst, E. 1992. Testing of fuzzy set theory in land suitability assessment for rainfed grain maize production. Pedologie. 42(2); 12947.

Triantafilis, J., and Mc Bratney, A.B. 1993. Application of continuous methods of soil classification and land suitability assessment in the lower Namoi valley. CSIRO Div. Soils, Melbourne, Div. Rep.,121: 172pp.

Towner, C.D. 1968. Variability of soil moisture in the black cracking clay soil of north-western New South Wales. Aust. J. Exp. Agric. Anim. Husb. 8:252-4

Trangmar, B. B., Yost, R. S. and Uehara, G. 1985. Application of Geostatistics to spatial studies of soil properties. Advances in Agronomy. 38:45-95.

Trangmar, B. B., Yost R. S. and Uehara G. 1987. Spatial variation of soil properties and rice yield on recently cleared land. J. Soil Soc. Am. 51(3):668-674.

USDA. 1982. Amendments to Soil Taxonomy. National Soil Taxonomy handbook 430-VI-Issue No 1. Soil Conservation Service. USDA, Washington, D.C.

Webster, R. 1968. Fundamental objections to the 7th approximation. J. of Soil Sci. 19:354-366.

Webster, R., and Burgess. T.M. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal Kriging. J. Soil Sci. 31: 505-524.

Webster, R, and Butler, B. 1976. Soil classification and survey studies at Ginninedera Aust. J. of Soil Research 14: 1-24.

Wilding, L.P. Dress, L.R. 1983. Spatial variability and pedology. p. 83-116. *In* L.P. Wilding, Smeck, N.E, and Hall, G.F. (ed.) Pedogenesis and soil taxonomy.I. Concepts and Interactions. Elsevier. Amstrdam.

Wilson, C.E. 1999. Soil conditions and farming practices affecting the yields and sustainability of maize (zea mays L.) crops in the Texcoco River Watershed, Central Mexico. Watershed Ecosystem M.Sc. Program, Trent University Peterborough, Ontario. Canada.

Zadeh L. A. , 1965 , Fuzzy sets. Information and Control 8: 338-353.

Appendix (1a): Field measurements for slope, depth and coarse fragments along with samples location.

Code	Easting (m)	Northing (m)	Slope (%)	Depth (cm)	Coarse Fragments %	Code	Easting (m)	Northing (m)	Slope (%)	Depth (cm)	Coarse Fragments %
1	517320	2156041	6	70	0	43	517393	2155125	18	40	0
2	517050	2155742	3	60	0	44	516156	2154387	18	40	3
3	522992	2152635	24	100	0	45	516633	2154343	8	40	3
4	521479	2152657	18	100	0	46	515638	2156847	4	30	1
5	521340	2152776	15	100	0	47	515925	2154329	8	35	1
6	521566	2153261	11	50	1	48	517878	2154824	7	90	0
7	521573	2153226	11	50	1	49	515830	2156609	7	90	0
8	520430	2154545	24	75	0	50	521624	2152612	6	60	0
9	520430	2154545	24	75	0	51	515114	2157211	6	35	0
10	516486	2155098	17	75	0	52	518618	2154446	5	35	6
11	521788	2152413	12	30	3	53	517430	2154855	27	30	0
12	521170	2153048	11	100	1	54	518483	2154551	6	35	0
13	520935	2152876	15	100	0	55	515678	2156503	7	100	4
14	520935	2152876	15	100	0	56	517396	2155135	12	35	0
15	520935	2152876	22	100	0	57	520174	2155229	8	20	3
16	521374	2152672	18	50	1	58	520000	2154874	6	35	9
17	521374	2152672	8	50	1	59	519736	2154463	4	40	3
18	516455	2153799	18	35	1	60	516761	2156316	9	40	1
19	521894	2153074	18	35	0	61	519978	2155103	11	35	6
20	520402	2154304	28	20	3	62	516664	2156412	15	50	3
21	521151	2539500	28	30	2	63	516734	2156097	5	30	3
22	521151	2159500	7	30	2	64	516566	2156343	11	35	2
23	516611	2154149	6	30	1	65	516677	2156320	6	40	0

24	516704	2154520	6	30	0	66	518979	2154493	8	35	0
25	517316	2154259	7	35	0	67	518944	2154393	8	50	0
26	516587	2154162	6	35	0	68	519794	2154039	9	40	0
27	516248	2154375	6	60	3	69	518741	2154656	6	30	0
28	516266	2154661	6	35	1	70	518586	2154718	10	35	0
29	516950	2154492	6	35	3	71	517069	2156047	5	20	1
30	516950	2154004	6	40	1	72	520679	2153832	5	40	1
31	515911	2154033	5	40	1	73	520881	2153587	8	30	0
32	517438	2153948	7	30	1	74	520617	2153326	8	60	1
33	516695	2154315	7	40	0	75	520183	2154082	7	30	1
34	516804	2154210	8	50	0	76	520838	2153633	11	60	1
35	516653	2154323	8	40	1	77	520158	2153757	11	40	3
36	516687	2154210	8	35	0	78	520994	2152722	21	40	10
37	516373	2157296	6	40	0	79	-	-	-	-	-
38	516738	2153972	6	40	0	80	-	-	-	-	-
39	516775	2154097	18	50	0	81	-	-	-	-	-
40	519691	2153744	27	30	0	82	-	-	-	-	-
41	519691	2153744	8	30	1	83	-	-	-	-	-
42	515911	2156467	8	90	0						

Appendix (1b): Soil chemical analysis

Code	pH	Organic Matter %	K meq/100g	Na meq/100g	Mg meq/100g	Ca meq/100g	CEC meq/100g	%Base Saturation	Exchange able Na %	EC mmhos/c
1	5.6	6.9	0.66	0.071	6.05	11.27	19.49	89	< 2	0.59
2	5.4	5.5	0.48	0.077	5.07	11.87	13.47	89	< 2	0.49
3	7.1	6.6	1.04	0.049	3.27	7.68	19.88	77	< 2	0.48
4	7.2	8.4	0.56	0.085	5.07	9.68	19.58	53	< 2	0.32
5	7.4	6.9	0.92	0.055	2.45	7.08	14.42	86	< 2	0.66
6	6.3	6.7	0.66	0.075	4.25	7.43	11.96	86	< 2	0.33
7	6.8	4.5	0.48	0.051	3.60	6.23	16.11	81	< 2	0.29
8	5.8	4.9	0.74	0.044	4.58	7.78	15.49	76	< 2	0.38
9	5.6	5.2	0.66	0.059	4.25	6.88	18.32	90	< 2	0.35
10	5.6	5.3	0.56	0.058	3.96	11.97	10.36	71	< 2	0.66
11	5.8	2.7	0.53	0.062	3.27	3.49	13.95	78	< 2	0.36
12	6.0	6.2	0.35	0.065	2.78	7.68	14.29	87	< 2	0.54
13	5.5	6.3	0.61	0.070	3.27	15.56	20.58	99	< 2	1.04
14	5.8	5.7	0.71	0.060	4.58	7.18	17.76	98	< 2	0.43
15	6.7	7.9	0.33	0.069	8.51	11.52	18.22	80	< 2	0.55
16	5.2	7.8	0.92	0.065	6.22	10.23	16.93	79	< 2	0.57
17	5.4	7.1	0.94	0.066	4.58	8.98	18.61	97	< 2	0.52
18	6.7	7.1	0.94	0.074	8.67	19.81	9.11	23	< 2	0.71
19	5.9	5.8	0.43	0.061	5.89	6.98	21.9	65	<2	0.45
20	6.0	7.0	0.92	0.067	7.86	9.33	19.68	73	< 2	0.62
21	7.1	3.5	0.025	0.062	0.49	1.59	26.05	91	< 2	0.87
22	7.3	7.3	1.48	0.070	5.73	6.48	25.73	91	< 2	0.57
23	6.3	5.8	0.79	0.051	4.91	9.83	19.54	87	< 2	0.40
24	6.6	6.2	0.56	0.058	5.89	12.57	14.69	95	< 2	0.37

25	7.0	7.1	0.66	0.062	8.18	14.52	19.69	93	< 2	0.71
26	5.7	6.0	0.46	0.054	5.40	12.02	15.97	91	< 2	0.38
27	5.7	4.5	0.20	0.045	3.76	8.88	19.67	82	< 2	0.29
28	5.9	5.7	0.69	0.054	5.89	12.12	22.73	86	< 2	0.68
29	6.1	6.9	1.43	0.044	7.20	15.41	20.65	99	< 2	0.60
30	6.2	4.8	0.38	0.059	4.74	9.68	15.72	862	< 2	0.34
31	6.3	5.6	0.63	0.056	6.05	11.22	19.65	95	< 2	0.43
32	5.9	6.3	0.99	0.063	5.89	11.82	26.16	98	< 2	0.69
33	5.5	5.9	0.48	0.066	5.89	11.47	21.66	97	< 2	0.46
34	6.0	5.6	1.91	0.051	6.38	12.62	16.88	91	< 2	0.66
35	6.0	5.5	0.23	0.072	5.07	10.23	13.9	68	< 2	0.38
36	5.9	5.6	0.48	0.056	5.56	10.92	19.27	57	< 2	0.35
37	5.8	5.7	1.86	0.057	4.91	12.37	21.66	78	< 2	0.55
38	6.5	6.6	0.51	0.050	5.89	14.82	30.69	82	< 2	0.53
39	6.3	5.8	0.53	0.066	5.24	10.67	16.29	93	< 2	0.48
40	5.8	4.5	0.48	0.040	4.24	7.78	24.25	90	< 2	0.39
41	6.1	7.1	1.32	0.054	3.93	7.88	25.01	84	< 2	0.50
42	5.5	6.0	0.40	0.049	6.22	5.68	24.58	93	< 2	0.50
43	5.8	6.3	0.53	0.053	5.56	17.81	17.21	65	< 2	0.96
44	6.6	7.1	1.20	0.070	8.84	9.88	24.3	82	< 2	0.44
45	5.9	4.9	0.86	0.065	3.11	9.33	19.00	92	< 2	0.54
46	6.1	5.5	0.33	0.55	5.89	16.41	30.05	64	< 2	0.49
47	6.3	5.6	0.81	0.071	7.69	11.32	22.2	83	< 2	0.51
48	5.9	6.6	2.89	0.055	5.56	12.37	11.2	68	< 2	0.28
49	5.9	5.7	0.48	0.039	4.74	10.77	21.77	99	< 2	0.83
50	6.3	4.4	0.33	0.060	3.93	11.47	18.4	99	< 2	0.69
51	5.4	4.9	0.56	0.063	4.74	10.32	10.05	100	< 2	0.44
52	5.8	7.0	1.17	0.045	9.00	17.71	22.43	95	< 2	0.58
53	5.5	5.2	0.38	0.050	4.91	16.96	36.77	96	< 2	0.65
54	5.3	5.9	1.27	0.055	4.74	8.28	30.31	77	< 2	0.24
55	4.6	2.8	0.25	0.063	1.31	7.53	15.48	91	< 2	0.33

56	5,7	8,2	0,86	0,058	1,80	12,17	16,82	85	< 2	0,73
57	6,0	5,9	0,40	0,063	5,24	12,67	15,75	91	< 2	0,38
58	6,3	4,1	0,46	0,052	3,76	5,86	14,89	100	< 2	0,36
59	6,2	5,1	0,51	0,060	3,43	14,82	16,46	93	< 2	0,38
60	6,1	4,8	0,43	0,063	5,07	15,02	29,5	60	< 2	0,75
61	5,5	6,7	0,69	0,076	3,73	16,06	31,18	91	< 2	1,08
62	7,3	9,6	3,75	0,065	8,35	22,75	19,16	96	< 2	0,41
63	6,3	9,5	1,40	0,071	9,66	24,40	14,39	82	< 2	0,76
64	6,2	7,8	1,76	0,042	5,24	22,10	15,98	99	< 2	0,24
65	5,8	4,7	0,30	0,066	3,11	8,48	22,36	93	< 2	0,53
66	6,0	10,2	2,07	0,053	7,20	26,54	23,26	85	< 2	0,28
67	5,4	4,7	0,40	0,054	5,56	9,28	17,07	92	< 2	0,28
68	5,9	4,9	0,40	0,052	2,12	10,82	9,48	82	< 2	0,29
69	4,2	5,2	0,33	0,056	4,25	8,98	6,67	-	-	0,37
70	5,3	6,0	0,69	0,050	9,66	10,81	-	-	-	0,42
71	6,3	12,3	1,04	0,087	8,84	17,51	-	-	-	0,03
72	5,0	7,0	0,38	0,065	5,89	12,62	-	-	-	0,40
73	6,8	9,5	0,69	0,064	7,36	9,48	-	-	-	1,20
74	7,5	11,3	1,76	0,063	6,71	18,76	-	-	-	1,03
75	6,7	6,8	0,28	0,039	3,27	10,23	-	-	-	0,29
76	6,2	6,4	0,20	0,072	4,25	8,58	-	-	-	0,59
77	6,7	15,2	1,37	0,066	12,28	24,60	-	-	-	0,44
78	6,5	9,9	0,89	0,053	9,00	12,57	-	-	-	0,60
79	6,2	9,4	0,63	0,056	6,05	14,92	-	-	-	0,31
80	5,7	6,4	0,56	0,062	5,07	8,88	-	-	-	0,21
81	5,3	11,0	0,61	0,058	8,67	14,37	-	-	-	0,27
82	6,4	6,0	0,46	0,040	2,29	5,98	-	-	-	-
83	5,3	7,7	-	-	-	-	-	-	-	-

Appendix (1c): Analysis of soil texture

Code	Clay %	Silt %	Sand %	Texture	Code	Clay %	Silt %	Sand %	Texture
1	12	32	56	S L	44	12	44	44	L
2	19	30	51	SL	45	13	35	52	s.L
3	17	43	40	SL	46	14	38	48	L
4	36	20	44	SL	47	15	54	31	L
5	19	33	48	L	48	23	46	31	L
6	20	25	55	L	49	19	36	45	s.L
7	15	25	60	SL	50	21	38	41	L
8	15	27	58	SL	51	28	53	19	SL
9	17	25	58	SL	52	27	41	32	SL
10	14	30	56	SL	53	11	32	57	LS
11	3	40	57	s.S	54	9	38	53	L
12	22	33	45	L	55	5	17	78	L
13	22	48	30	L	56	19	70	11	L
14	23	48	29	s.L	57	22	41	37	L
15	20	9	71	SL	58	17	35	48	L
16	37	15	48	CL	59	18	35	47	SCL
17	30	46	24	SL	60	15	32	53	L
18	18	33	49	L	61	31	54	15	L
19	16	33	51	L	62	16	37	47	L
20	13	36	51	L	63	22	44	34	L
21	30	15	55	SL	64	13	33	54	sL
22	13	29	58	SL	65	17	43	40	L
23	24	40	36	L	66	28	47	25	CL
24	23	41	36	L	67	27	47	26	L
25	34	54	12	SCL	68	30	46	24	L
26	31	52	17	SCL	69	16	36	48	sS
27	9	28	63	s.S	70	20	43	37	SL

28	15	34	51	L	71	8	31	61	sS
29	23	50	27	L	72	10	31	59	SL
30	20	38	42	L	73	7	32	61	L
31	16	35	49	s.L	74	10	35	55	LS
23	26	42	32	L	75	23	47	30	LS
33	22	54	24	SL	76	1	21	78	L
34	16	38	46	L	77	5	16	79	sS
35	13	31	56	s.S	78	24	47	29	sS
36	14	34	52	SL	79	6	37	57	L
37	8	28	64	L	80	2	30	68	sS
38	18	28	54	SL	81	26	30	44	sS
39	18	44	38	L	82	7	22	71	
40	10	29	61	SL	83	5	23	72	
41	20	40	40	s					
42	10	30	60	L					
43	9	27	64	L					

S= Sand s= Silt C= Clay L= Loam

Appendix (2): Meteorological stations data

No.	Station	Easting (m)	Northing (m)	Length of Growing Period (days)	Annual Rainfall (mm)	Rainfall Growing Season (mm)	Mean Temperature growing Season (°C)	Mean Minimum Temperature growing season (°C)
1	Atenco	508800	2161600	155	586.84	529.99	16.2	8.4
2	Eltejcote	510250	2150400	161	712.43	658.64	12.1	7.3
3	La Grande	512300	2161500	163	596.25	539.94	15.9	7.7
4	San Andres	513300	2161300	163	548.52	501.22	16	8.2
5	Santa Maria	517750	2155150	163	914.29	839.35	13.5	7.3
6	Texcoco	512200	2157900		545.97	496	17.4	9
7	Chapingo	512150	2156000		523.45	475	16.4	9
8	Montecillio	509500	2151450		627.72	573	16	9
9	Lomas De Cristo	513100	2152150		613.89	560	16	9
10	Purificacion	519850	2158950		569.20	518	16	9
11	Tlaixpan	519500	2157850		561.75	511	16	5
12	Tequexquinahuac	518400	2152900					

Appendix (3a): Landscape and Soil requirements for Maize

Land Characteristics	Land class degree of limitation and rating scale					
	S1		S2	S3	N1	N2
Topography	0	1	2	3		
Slope (%)	0-2	2-4	4-8	8-16	30-50	≻ 50
Wetness						
Flooding	Fo (no flood)	Fo (no flood)	Fo (no flood)	F1 (slight)	F1	F2 (moderate)
Drainage	good	Moderate	Imperfect	Poor	Poor	very poor
Physical Soil Characteristics						
Texture	C-60s, SiCs, Co, SiCL, CL, Si, SiL	C+60sv, C-60v SC, L, SCL	C+60v, SL, LfS, LS	LcS, f S	-	Cm, S, cS
Coarse fragments (Vol. %)	< 3	3-15	15-35	35-55	-	≻ 55
Soil Depth (cm)	>100	75-100	50-75	20-50	-	< 20
Fertility Characteristics						
CEC (meq/100)	> 24	16-24	< 16	< 16		
Organic matter %	>2	1.2-2	0.8-1.2	< 0.8		
Base Saturation	> 50	35-50	20-35	< 20		
Salinity and Alkalinity						
EC, mmhos/cm	0-2	2-4	4-6	6-8	8-12	≻ 12
ESP	0-8	8-15	15-20	20-25	-	≻ 25
pH	5.5-8	5.2-5.5		4.5-5	4-4.5	8.5-10.5

Glossary of the texture symbols:

Cm: massive clay

SiCm: massive silty clay

c+ 60,v : fine clay, vertical structure

c+ 60, s: fine clay, blocky structure

c- 60, s: clay, blocky structure

SiCs; silty clay, blocky structure

Co: clay, ocisol structure

SiCL: silty clay loam

CL: clay loam

Si: Silt

SiL: silt loam

SC: sandy clay

L: loam

SCL: sandy clay loam

SL: sandy loam

LfS: loamy fine sand

LS: loamy sand

LcS: loamy coarse sand

fS: fine Sand

S: sand

cS: coarse sand

Appendix (3b): Climatic requirements

Climatic Characteristics	Climatic class degree of limitation and rating scale					
	S1		S2	S3	N1	N2
	0	1	2	3		
Annual Rainfall (mm)	850-1250	750-850	600-750	500-600	-	< 500
Length Growing Season (days)	150-220	130-150	110-130	90-110	-	<90
Rainfall Growing Season (mm)	800-1200	700-800	600-700	500-600	-	< 500
Mean Temperature (^{0}C) Growing Season	22-26	18-22	16-18	14-16	-	<14
Mean Minimum Temp. Growing Season (^{0}C)	16-18	12-16	9-12	7-9	-	<7
Relative Humidity Devel. Stage %	50-80	42-50	36-42	30-36	-	<30
Relative Humidity Maturation Stage%	30-50	24-30	20-24	<20		
n/N Devel. Stage	0.5-0.6	0.35-0.5	<0.35			
n/N Maturation Stage	>0.7	0.7-0.5	<0.5			

n/N Devel. Stage=relative humidity development stage
n/N Maturation. Stage= relative humidity maturation stage

Appendix (4): Climatic requirements decision tree.

Severity Level decision tree for LUT,LUR 'MA,c'
AR_annual rainfall__3_ml (moderately low) [500-600 mm]________________________
_LGS_Length Growing Season__2_c2 (second) [90-110 days]______________________
__RGS_Rainfall growing season__3_l (low) [500-600 mm]______________________
___MGS_Mean temperature growing season__3_l (low) [14-16 °c]________________
____MMTG_Mean minimum temperature grow. season__3_ml (moderately low) [7-9 °c]_
_____RH_Relative humidity devel.stage__3_ml (moderately low) [30-36 %]__________
______RM_Relative humidity maturation stage__1_l (very low) [0-20 %]___________
_______IRDS_Insolation Rate Development__1_3d (third) [0-.35 rate]____________
________IRMS_Insolation Rate Maturation Stage_____________________________
__________1_l (low) [0-.5 rate]________* 4 (S3)__________________________
__________2_h (high) [.5-.7 rate]_______=1______________________________
__________3_vh (very high) [.7-1 rate]__=1______________________________
__________?_[???]______________________?________________________________

Appendix (5a): Current paradigm database

Polygon ID	AR	LGP	RGS	MTGS	MMTGS	RHDS	RHMS	IRDS	IRMS	SLOPE	FLOOD	DRAINAGE	SOIL DEPTH	CEC	BS	OM	pH	ESP	EC	TEXTURE CLASS
1	750	166	650	16	8	80	50	0.6	0.7	15	0	MD	15	13.5	100	1.8	7.1	15	<2	SL
2	750	166	650	16	8	-	-	-	-	15	0	GD	100	34.5	100	5.6	10.4	15	20	CL
3	700	163	540	16	8.5	-	-	-	-	15	0	GD	60	28	100	2.4	8.1	15	2.0	CL
4	775	166	650	14.5	8	-	-	-	-	40	0	GD	60	28	100	2.4	8.1	15	2.0	CL
5	900	168	610	14.5	6	-	-	-	-	40	0	GD	60	34.5	100	5.6	10.4	15	20	CL
6	800	165	620	14.5	6.5	-	-	-	-	40	0	GD	19	28.5	100	2.7	10.4		20	CL
7	750	167	620	14.5	6.5	-	-	-	-	40	0	GD	60	28	100	2.4	8.1		20	CL
8	800	168	620	14.5	7	-	-	-	-	40	0	GD	100	34.5	100	5.6	10.4		20	CL
9	800	166	620	14.5	7	-	-	-	-	40	0	GD	60	28	100	2.4	8.1			CL
10	800	166	600	14.5	7	-	-	-	-	40	0	GD	100	34.5	100	5.6	10.4			CL

AR=ANNUAL RAINFALL, LGP=LENGTH OF GROWING PERIOD, MTGS=MEAN TEMP. GROWING SEASON, MMTGS=MINIMUM MEAN TEMP. GROWING SEASON, RHDS=RELATIVE HUMIDITY DEVELOPING STAGE, RHMS= RELATIVE HUMIDITY MATURATION STAGE, BS= BASE SATURATION OM=ORGANIC MATTER, ESP=EXCHANGABLE SODIUM PERCENTAGE, GD=GOOD, 0=NO FLOOD

APPENDIX (5b): Proposed Paradigm database

Site No.	AR	LGP	RGS	MTGS	MMTGS	RHDS	RHMS	IRDS	IRMS	SLOPE	FLOOD	SOIL DEPTH	CEC	BS	OM	pH	ESP	EC	TEXTURE
1	700	166	640	16	7.5	80	30	0.5	0.7	3	0	60	19.4	89	8.4	5.4	0.39	0.49	SL
2	660	163	580	15.5	7.5	80	30	0.5	0.7	18	0	100	19.8	77	6.9	7.2	0.43	0.32	C.S
3	660	163	580	15.5	7.5	80	30	0.5	0.7	15	0	100	19.5	53	6.7	7.4	0.23	0.66	SL
4	660	163	580	15.5	7.5	80	30	0.5	0.7	11	0	50	14.4	86	4.5	6.3	0.52	0.31	SL
5	680	163	580	15.5	7.5	80	30	0.5	0.7	11	0	50	11.9	86	4.9	6.8	0.42	0.29	SL
6	680	164	600	15.5	7.5	80	30	0.5	0.7	24	0	75	16.1	81	5.2	5.8	0.27	0.38	SL
7	700	164	600	16	7.5	80	30	0.5	0.7	24	0	75	15.4	76	5.3	5.6	0.38	0.35	SL
8	660	165	620	16	7.5	80	30	0.5	0.7	17	0	75	18.3	90	2.7	5.6	0.31	0.66	C.s
9	660	163	580	16	7.5	80	30	0.5	0.7	12	0	130	10.3	71	6.2	5.8	0.60	0.36	C.s
10	660	163	580	16	7.5	80	30	0.5	0.7	11	0	100	13.9	78	5.7	6.0	0.47	0.54	C.s
11	660	163	580	16	7.5	80	30	0.5	0.7	16	0	100	14.2	87	7.9	5.8	0.42	0.43	SL
12	660	163	580	16	7.5	80	30	0.5	0.7	15	0	100	20.5	99	7.8	6.7	0.33	0.55	C.s
13	660	163	580	15.5	7.5	80	30	0.5	0.7	22	0	50	17.7	99	7.1	5.2	0.36	0.57	C.s
14	660	163	580	15.5	7.5	80	30	0.5	0.7	18	0	50	18.2	98	5.8	5.4	0.36	0.52	C.s
15	660	164	580	15.5	7.5	80	30	0.5	0.7	18	0	35	16.9	80	7.0	5.9	0.47	0.45	C.s
16	660	164	600	15.5	7.5	80	30	0.5	0.7	18	0	20	18.6	79	3.5	6.0	0.69	0.62	C.s
17	660	164	620	15	7.5	80	30	0.5	0.7	28	0	30	9.1	97	7.3	7.1	0.58	0.87	SL
18	660	164	620	15	7.5	80	30	0.5	0.7	28	0	30	21.9	23	5.8	7.3	0.28	0.57	C.s
19	700	165	620	15	7.5	80	30	0.5	0.7	7	0	30	19.6	65	6.2	6.3	0.22	0.40	C.s
20	700	165	640	15	7.5	80	30	0.5	0.7	6	0	30	26.0	79	7.1	6.4	0.24	0.37	C.s
21	700	165	640	15	7.5	80	30	0.5	0.7	6	0	35	25.7	73	6.0	6.6	0.28	0.71	C.s
22	700	165	620	15	7.5	80	30	0.5	0.7	7	0	35	19.5	91	4.5	7.0	0.30	0.38	C.s
23	700	165	600	15	7.5	80	30	0.5	0.7	6	0	60	14.6	91	5.7	5.7	0.27	0.29	C.s
24	700	165	620	15	7.5	80	30	0.5	0.7	6	0	35	19.6	87	5.6	5.7	0.28	0.68	C.s

25	700	164	600	15	7.5	80	30	0.5	0.7	6	0	40	19.6	91	5.6	6.2	0.28	0.43	C.s
26	700	165	620	15	7.5	80	30	0.5	0.7	5	0	30	22.7	82	6.3	6.3	0.28	0.69	C.s
27	700	165	620	15	7.5	80	30	0.5	0.7	7	0	40	20.6	86	5.9	5.9	0.32	0.46	C.S
28	700	165	620	15	9.0	80	30	0.5	0.7	8	0	40	15.7	99	5.5	5.5	0.46	0.38	C.s
29	700	165	620	15	7.5	80	30	0.5	0.7	8	0	35	19.6	86	5.6	6.0	0.28	0.35	SL
30	640	164	580	15	7.5	80	30	0.5	0.7	8	0	40	26.1	95	5.7	6.0	0.28	0.55	SL
31	700	165	620	15	7.5	80	30	0.5	0.7	8	0	40	21.6	80	6.6	5.9	0.23	0.53	C.S
32	700	165	620	16	7.5	80	30	0.5	0.7	6	0	50	16.8	97	5.8	5.8	0.29	0.48	SL
33	660	164	580	16	7.5	80	30	0.5	0.7	18	0	30	13.9	91	4.5	6.5	0.29	0.39	C.s
34	660	163	580	15	7.5	80	30	0.5	0.7	27	0	30	19.2	66	7.1	6.3	0.28	0.50	C.s
35	700	166	640	15.5	6.5	80	30	0.5	0.7	8	0	90	21.6	57	6.0	5.8	0.22	0.5	SL
36	700	165	620	15.5	6.5	80	30	0.5	0.7	8	0	40	30.6	78	6.3	6.1	0.17	0.96	C.s
37	660	165	600	14	7.5	80	30	0.5	0.7	18	0	40	16.2	82	4.9	5.8	0.4	0.54	SL
38	660	165	600	15	8.0	80	30	0.5	0.7	8	0	30	24.5	93	5.5	6.6	0.22	0.49	C.s
39	780	168	700	15	7.5	80	30	0.5	0.7	4	0	35	25.0	90	5.6	5.9	0.23	0.51	C.S
40	700	165	600	16	9.0	80	30	0.5	0.7	8	0	90	24.5	84	6.6	6.1	0.22	0.28	C.s
41	700	165	620	14.5	7.0	80	30	0.5	0.7	7	0	60	24.3	65	4.4	5.9	0.25	0.69	C.s
42	660	164	580	16	9.0	80	30	0.5	0.7	8	0	35	19.0	82	4.9	5.9	0.15	0.44	C.S
43	780	168	700	16	9.0	80	30	0.5	0.7	6	0	35	30.0	92	7.0	6.3	0.24	0.58	SL
44	680	164	580	14	7.5	80	30	0.5	0.7	27	0	35	22.2	64	5.9	5.8	0.57	0.24	SL
45	640	163	580	15	7.0	80	30	0.5	0.7	6	0	100	11.2	83	2.8	5.5	0.26	0.33	LS
46	780	168	700	16	9.0	80	30	0.5	0.7	7	0	35	21.7	68	8.2	5.3	0.34	0.73	C.S
47	720	166	640	14	7.5	80	30	0.5	0.7	6	0	20	18.4	99	5.9	4.6	0.52	0.35	C.s
48	660	164	580	15.5	8.0	80	30	0.5	0.7	12	0	35	10.0	99	4.1	5.7	0.34	0.38	SL
49	780	168	700	15.5	8.0	80	30	0.5	0.7	4	0	35	22.4	90	6.7	6.2	0.17	0.75	C.S
50	660	165	600	15.5	8.0	80	30	0.5	0.7	9	0	50	36.7	95	9.6	6.1	0.13	1.08	C.S
51	680	163	600	15.5	8.0	80	30	0.5	0.7	15	0	35	30.3	96	7.8	7.3	0.40	0.76	C.S
52	680	165	600	15.5	8.0	80	30	0.5	0.7	5	0	40	15.4	77	4.7	6.3	0.32	0.24	C.s
53	680	165	600	15	7.0	80	30	0.5	0.7	6	0	50	16.8	91	4.7	5.8	0.33	0.28	C.s

54	680	163	600	15	7.0	80	30	0.5	0.7	8	0	40	15.7	85	4.9	6.0	0.37	0.28	SL
55	700	165	620	15.5	7.0	80	30	0.5	0.7	8	0	30	14.8	91	5.2	5.4	0.30	0.29	SL
56	700	165	620	15	7.0	80	30	0.5	0.7	8	0	35	16.4	99	6.0	5.9	0.22	0.37	SL
57	680	165	600	15	7.0	80	30	0.5	0.7	9	0	20	29.5	93	12	4.2	0.20	0.42	C.s
58	700	166	640	15	7.0	80	30	0.5	0.7	6	0	40	31.1	60	7.0	5.3	0.33	0.03	LS
59	720	166	640	15.5	7.5	80	30	0.5	0.7	10	0	30	19.1	91	9.5	6.3	0.27	0.40	C.s
60	700	165	640	15.5	7.0	80	30	0.5	0.7	5	0	30	14.3	96	6.8	6.8	0.45	1.03	C.S
61	680	164	580	15.5	7.0	80	30	0.5	0.7	8	0	60	15.9	82	6.4	7.5	0.23	0.29	C.S
62	660	164	580	15.5	7.0	80	30	0.5	0.7	7	0	40	22.3	99	9.9	6.2	0.24	0.44	LS
63	680	164	600	16	7.0	80	30	0.5	0.7	11	0	40	23.2	93	9.4	6.7	0.36	0.60	SL
64	680	164	600	15	7.5	80	30	0.5	0.7	11	0	20	17.0	85	8.4	6.5	0.42	0.31	SL
65	680	164	600	16	7.0	80	30	0.5	0.7	21	0	25	9.4	92	6.0	5.7	0.33	0.21	SL
66	680	164	600	16	7.0	80	30	0.5	0.7	40	0	55	6.6	82	7.7	5.3	0.78	0.27	SL