A KNOWLEDGE BASED APPROACH TO THE MANAGEMENT OF SOIL EROSION INFORMATION IN DEVELOPING COUNTRIES.

E.G. Mtalo and E Derenyi

Department of Surveying Engineering
University of New Brunswick
Fredericton, N.B. CANADA

COMMISSION VII

ABSTRACT

The study and control of soil erosion and the associated environmental degradation is a difficult problem because it involves a very wide body of knowledge. The use of knowledge based expert systems in this domain may improve the current efforts by providing an efficient tool for inter-disciplinary technology transfer and application. Domain expert systems may also be used as substitutes for human experts to provide guidance to less skilled soil erosion domain personnel.

This paper describes the prototype knowledge based system Soil Loss Estimation and Modelling System (SLEMS) and shows how such a system can be used to solve some soil erosion problems. This system was developed and implemented at UNBS on a Sun workstation for the management and application of remote sensing and geographic information in soil loss estimation and modelling problems.

Key Words: Knowledge Base, Soil Erosion, Environment, GIS, Remote Sensing, Developing Countries

1. INTRODUCTION.

Soil erosion is defined as the process whereby detachment and transportation of soil from its natural location takes place usually with adverse impact on the environment. Factors and causes of soil erosion are characterized as: climatic and hydrologic agents such as rain, runoff, antecedent moisture, wind, geographical position, and altitude; morphological agents including slope, slope length and surface roughness; geological and soil agents which include soil type, parent material and soil texture; vegetative agents such as plant cover type, density and height; technical agents consisting of conservation management, tillage systems, farm equipment and construction activities; and social economic agents such as population distribution and distribution, and other human activities(Wischmeier and Smith, 1957, Goldman et al, 1986). Morgan (1986) considers soil erosion to be a multi-faceted problem whose solution involves five issues: policy, measurement and inventory of soil erosion extent and severity, assessment and evaluation of soil loss conservation efforts, modelling and estimation of soil loss and social and economic issues.

The five components of the soil erosion management problem are strongly interrelated. The various factors influencing soil erosion are also intricately intertwined. Clearly soil erosion is a highly complex problem whose solution requires harnessing and integrating information and technology from a wide range of disciplines. Moreover a considerable amount of the information needed to model and estimate soil erosion is fuzzy or non-precise. At the same time soil erosion evaluation often relies on fuzzy decision.

Monitoring, management, and control of soil erosion in developing countries is a critical issue because the economic base of most developing countries is mainly agricultural. The dangers posed by increased soil erosion to the economies of developing countries are therefore very great.

In the developed world it has become fashionable to handle complex problems which require intensive human expertise by knowledge based computer systems. This approach extends the human ability to analyze complex systems in order to provide accurate and timely answers to difficult problems. The expert systems technology releases the human expert from tiresome, repetitive procedures so that s/he may devote more time on designing efficient problem solving strategies. One area where the knowledge based systems approach may be beneficial is in the solution of soil erosion problems. However at present there have been very few expert systems designed to solve soil erosion problems.

In developing countries, where lack of human expertise in all fields is often acute, knowledge based systems in the form of expert systems could become an important tool for technology transfer and application. Domain expert systems can be used as substitutes for human experts in providing guidance to less skilled personnel in developing countries.

Assuming an appropriate framework in which local experts work together with foreign experts to develop knowledge based soil erosion applications the benefits which may be realised include:

(i) Efficient and sustainable transfer of monitoring and control technology,
(ii) On the job instruction on problem solving procedures by domain expert systems allowing less skilled technicians to perform tasks requiring expert knowledge,
(iii) Systematic training of local technicians and skilled labour on the use and maintenance imported technology,
(iv) Reduced dependence on foreign expatriates and boosting up self-reliance.

This paper discusses an experimental knowledge based system (SLEMS), developed as part of an academic research, for modeling and estimating soil erosion (Mtalo, 1990). Because of the need for affordable technology in developing countries the system was developed with a view to making it locally implementable. The system is therefore designed to provide a knowledge based management of attributive information and data required for modelling and estimating soil erosion.

1.1 Knowledge Representation Issues.

Knowledge based systems, differ from conventional software systems by the manner in which the knowledge required to solve specific problems is organised. In Knowledge based systems the facts (data) and knowledge needed for problem solving are extracted and stored in a knowledge base using special knowledge representation structures. The knowledge may be stored as a set of if...then... rules, frames and semantic networks (Firebaugh, 1988; Barr et al, 1989; Luger and Stubblefield, 1989). The process of extracting information from the knowledge base is referred to as knowledge inference. By using deductive and inductive inference techniques, knowledge not directly stored in the knowledge base can be inferred.

Systems which have knowledge about a specific and well defined domain of expertise, are called expert systems. The introduction of expert systems transformed artificial intelligence from a purely research domain into an application field. MYCIN and the INTERNIST were among the earliest practically useful rule based expert systems (Barr and Feigenbaum, 1981,1982).

Among the few expert systems designed for solving soil erosion problems is PLANTING commissioned by the United States
Department of Agriculture Soil Conservation Service (USDA-SCS). The system, which is based on the EXSYS shell, uses a knowledge base of farm equipment and soil characteristics to assist farmers in selecting the best combination of farm equipment and conservation planting technology to minimize soil erosion (Morrison et al, 1989).

Domain independent expert systems which contain only search control and general inference procedures are referred to as expert shells. Expert shells can be used to create expert systems for specific domains of applications. Depending on the knowledge representation method used expert shells may be characterized as rule based or frame based. Based on the search control or inference strategy used expert shells may also be characterized as forward chaining, backward chaining, or forward and backward chaining (Barr and Feigenbaum, 1981, 1982).

2. THE SLEMS CONCEPT.

SLEMS is essentially a rule based and semantic network based expert system. Knowledge is therefore captured and stored in the SLEMS knowledge base in the form of either a semantic network of facts or rules. Fuzzy or vague information is stored directly in the form of fuzzy expressions which can be queried by a fuzzy query processor as explained later.

The capture and representation of domain knowledge is, generally, an iterative two step process. First a knowledge engineer supported by the domain expert analyzes and abstracts the structure of the domain knowledge. The knowledge engineer then stores the abstracted knowledge using appropriate knowledge representation such as frames, semantic networks or rules. After knowledge acquisition the domain expert performs validation and verification tests on the system. The two processes are repeated as necessary to refine the representation of the domain knowledge (Ebrahimi, 1987 and Green and Keys, 1987).

The SLEMS uses both the rule based representation and semantic network structures to capture domain knowledge. The rules are designed as simple English-like "IF ... THEN ..." rules. These are compiled interactively by the SLEMS rule editor in the form of premises and conclusions. However in one of the SLEMS sub-system, LEARN, real world objects and facts are represented by a semantic network of object triplets. The SLEMS knowledge base is queried by a layer of specialized inference operators which exploit the hierarchical structure of the semantic network to infer facts not directly stored in the knowledge base. Alternatively a backward chaining strategy is used to search for knowledge base objects satisfying the premises of a rule. When a rule is satisfied the rule conclusion is asserted as a new fact and stored into the knowledge base.

The design of the SLEMS enables the capture and storage of three kinds of knowledge: knowledge directly needed for the solution of soil erosion problems, auxiliary knowledge required for the extraction of soil erosion information from external sources, such as, aerial photointerpretation and remote sensing, and knowledge about the organization and use of the SLEMS knowledge base.

The soil erosion domain knowledge stored in the SLEMS is based on Wischmeier's empirical formulae called the Universal Soil Loss Equation (USLE). It includes the six parameters, S, L, K, C, R, P, which represent the erosive effects of terrain morphology, soil characteristics, vegetation cover, hydrological and meteorological factors, and conservation practice (Wischmeier, 1984; Meyer, 1984). The USLE is widely used by soil conservation and agricultural experts as a design tool for soil erosion control.

Several variations of the USLE exist, each designed to handle different geographic, topographic, hydrologic, soil and ground cover conditions. The knowledge required by a soil erosion domain expert to select specific USLE models and compute values for the model parameters can be analyzed and organized by semantic network representation. For example Figure 1 shows a semantic network representation of the USLE concept. In this scheme the top of the hierarchy contains a general USLE model characterized by a default parameter set (PARAMETERS) whose instances are the six USLE parameters K, S, C, R, L, P. Specific USLE models, such as for example the USLE_1 which might represent the USLE model for Eastern Canada, appear lower in the semantic network hierarchy. Specific models are characterized by parameter sets containing both specific and default parameter descriptions. Each parameter instance is characterized by a value (VAL.). The semantic network organization thus enables missing parameters to be inherited from models defined higher up in the hierarchy.

Soil erosion is a complex problem. The study and the solution of soil erosion problems requires complex multi-variable information and a multi-disciplinary approach to information management and analysis. Certain aspects of the problem such as data capture and processing, information extraction, and knowledge acquisition and extraction are amenable to solution by computerized knowledge intensive methods. Figure 2 is a functional model of the soil loss estimation and modelling problem. Thick arrows in the figure indicate the forward flow of information, from the data capture stage to the knowledge extraction and application stage. Thin arrows indicate the interplay between the various components of the problem.

![Figure 1: A Semantic Network Representation of USLE Concepts](Source: E.G. Mtalo, 1990, page 65).
3. SYSTEM DESIGN AND COMPONENTS OF SLEMS.

SLEMS is an experimental knowledge based system prototype designed and developed for application in soil loss estimation and modelling as part of an Msc. thesis research at the UNB CanLab-INSPIRE in 1990 (Mtalo, 1990). At the current level of development the SLEMS can only manipulate attributive data and information. SLEMS consists of the following sub-systems (Figure 3):

- The EXPERT,
- The LEARN,
- The FUZZ,
- The Data Base Management System.

The first three sub-systems perform knowledge based functions. The EXPERT and the LEARN sub-systems are based on algorithms and source code published by Schildt (1987), modified and augmented for SLEMS implementation (Mtalo, 1990). The FUZZ sub-system was developed to process fuzzy knowledge. The Data Base Management System based on the CDATA (Stevens, 1987) performs conventional database management functions for the system. The inference layer consists of operators which exploit the semantic links between knowledge base objects to facilitate extraction of facts not directly stored in the knowledge base. The low level operators are basic search operators.

All the system modules were written in the C programming language and implemented on a SUN 4 Work Station. Since then a demonstration version of the SLEMS EXPERT was implemented for DOS based PC. Currently a new DOS version based on object oriented programming is being developed using BORLAND C++. In the new version the CDATA interface will be replaced by a new one based on the Paradox Engine. As mentioned in the introduction the primary objective is to produce a system which is affordable by developing countries.

3.1 The Rule Based EXPERT

The EXPERT is a rule based expert shell whose main function is to capture procedural and classification knowledge in the form of English-like "IF... THEN..." rules. Given certain facts the EXPERT uses a backward chaining search strategy to locate a rule satisfied by the given facts. If no rule is directly satisfied by the given facts the inference mechanism tries to find if there is a rule whose conclusion would provide the additional facts needed to satisfy another rule. Newly deduced facts are added into the knowledge base and used in subsequent inference.

At a preliminary stage the knowledge required to solve specific tasks must first be analyzed and organised into a decision tree where branches indicate alternate solution paths or related components of the knowledge. There are three objectives to the preliminary analysis. The first and most important objective is to enable the domain knowledge to be broken down into smaller manageable chunks without losing the semantic relationships inherent within the body of knowledge. Secondly the resulting structure (Figures 4) facilitates easy search of the resulting knowledge base. Thirdly the organization of the domain knowledge into a semantic network structure enables the direct translation of the structure into "IF...THEN..." rules and facilitates checks against errors in the generation of rules.

The structure of the knowledge required to calculate soil loss by the USLE model is shown in Figure 4. Based on this the first stage in the estimation of soil loss involves the selection of a specific model. Using the decision tree model USLE models may be classified by geographic area (e.g. Eastern Canada), by inventor (e.g. Wischmeier or Holy) or by method, such as, MUSLE (modified USLE), SLEMSA (modified soil estimation model for Africa), Revised Slope Factor, or Varying K (varying soil erodibility factor) as shown in Figure 4.

If, for example, the modified USLE (MUSLE) is selected then...
the value of the runoff curve number (CN) for the particular area must first be determined. The CN is an empirical measure of the runoff potential of a region which together with two parameters $S', P'$ (functions of the basin retention potential and precipitation) is used to calculate the potential runoff ($Q$) and soil loss $A(Q, S', L, K, C, P)$ (Bondelid, T. R. et al., 1980). If the CN values are not available default values must be estimated. This, in turn, requires the choice of a specific method such as the US Department of Agriculture method (USDA AMC-II CN METHOD), the US Geological Surveys method (USGS LUDA CN METHOD) or from Landsat data (LANDSAT CN METHOD) as indicated in Figure 4.

When a specific model has been chosen parameter values must be determined. Missing parameters must be computed or adopted from default values. This in turn requires further choice of a parameter estimation method (USLE PRMS in Figure 4) which depends on the required parameter and USLE model chosen. If, for example, the value of the slope factor $S$ is missing and the terrain slope is known (e.g. from topographic maps or digital terrain models) different formulae must be used to estimate its value depending on the magnitude of the slope gradient and the length of the slope (Figure 4). The estimated value is then plugged into the appropriate USLE model (e.g. $A(R, S, K, L, C, P)$) to compute the soil loss.

Figure 5 illustrates how specific knowledge on the selection of default C factors for the Universal Soil Loss Model (USLE) can be organised by method and location. Methods can be classified by inventor (e.g. HOLY and WISCHMEIER) or proprietor (e.g. USDA). Within each subcategory (e.g. HOLY) default parameters may be selected according to the crop growth period (PERIOD 1, PERIOD 2 ... etc). Where location specific parameters are available (e.g. GRANDFALLS) default C factors may be organised according to crop type (e.g. POTATO, GRAIN and HAY or POTATO and BROCCOLI). The selection of default parameters within each crop type may then be done according to the Crop Rotation Cycle such as P/P/P/H indicating three potato seasons followed by grain and hay rotations (Figure 5).

The solution of soil erosion problems requires inputs from a wide range of sources. In particular knowledge about ground cover (e.g. crop type and crop rotation cycle) is essential in estimating the C factor, which is, a measure of the degree of protection offered by ground cover against rain (and wind) erosion. Plant cover parameters, such as cover type and density, can be assessed from aerial photographs by aerial photointerpretation or by digital image analysis techniques. The knowledge needed to extract such information can be analyzed and organised into a semantic network of related knowledge. Figure 6 shows the decision tree resulting from the application of this strategy to the method of dichotomous photointerpretation of crop cover from aerial photographs. In this case each node in the tree represents a binary decision where the photo interpreter must make a choice between two alternatives as indicated on the associated key (Figure 6). The terminal nodes indicate the required classes. SLEMS can therefore be used to assist in the visual interpretation of ground cover based on the stored aerial photointerpretation knowledge.

In order to compile knowledge on a particular subject, the domain knowledge must first be analyzed and organised as explained. The resulting decision tree must then be translated into "IF... THEN..." rules which are then compiled and stored by the system's rule editor. Using the SLEMS EXPERT a rule base containing: rules for the selection of soil erosion models and their parameters, soil taxonomic classification rules, rules for the selection of spectral bands and their application to the classification of ground features, and the procedure for dichotomous air-photo interpretation was successfully compiled.

The system can also be used to record knowledge on the systems internal organization. The rules resulting from this are referred to as meta rules and they contain knowledge about the system and its application. Meta rules provide guidance to the user in the application of the knowledge contained in the knowledge base. They can also provide guidance in the compilation of knowledge.

A demo prototype of SLEMS was introduced to a group of multi-disciplinary local experts at the Ardhi Institute in Tanzania who were then asked to comment on the viability of the expert
systems technology. Their response was generally favorable and they pointed out the need to carry out a more exhaustive investigation for the purposes of refining the prototype.

3.2 A Structured Approach to the Compilation of Data and Facts.

The LEARN sub-system is an expert shell which captures and organizes knowledge in the form of a semantic network of English-like subject-verb-object triplets. Knowledge representation using the LEARN sub-system involves the analysis and restructuring of domain knowledge into Subject-Verb-Object (SVO) triplets where, “Subject”, represents a domain concept, “Verb”, represents a semantic relationship and, “Object”, represents an attribute or property of the “Subject”. If the “Object” part is a phrase assigning value to an attribute of the object then the SVO triplet is equivalent to an “Object-Attribute-Value” (OAV) triplet.

The query processor of the LEARN sub-system consists of several specialized inference operators, collectively, referred to as the inference layer (Figure 3). These operators exploit semantic links among the SVO object triplets to determine class relationships and extract the attribute values of the stored triplets. A section of the LEARN sub-system operators and their functional characteristics are shown in Figure 7 where S, V, O, SV, SO, VO and SVO are basic search operators, (s,v,o) represents the search space (database) and (S), (V), (O) and (True, False) represent solution spaces.
CROP CLASSES:
AG: START WITH AGRICULTURAL LANDS.
AG1: NON PRODUCTIVE LANDS.
AG211: TIMBERLAND.
AG212: GRASSLANDS.
AG22: FALLOW LANDS.
AG221: ORCHARDS.
AG222: VINE AND BUSH CROPS.
AG2221: ROW CROPS.
AG2222: IRRIGATED PASTURE LANDS.
AG22222: CONTINUOUS COVER CROPS, SMALL GRAINS, HAY ETC.

CLUES FOR INTERPRETATION:
AG1: VEGETATION & SOIL ABSENT OR OBSCURED BY ROCK ETC.
AG2: VEGETATION CLEARLY DISCERNIBLE ON PHOTOGRAPH.
AG21: CULTIVATION PATTERN ABSENT; FIELD BOUNDARIES IRREGULARLY SHAPED.
AG22: CULTIVATION PATTERN ABSENT; FIELD BOUNDARIES REGULARLY SHAPED.
AG211: TREES PRESENT, COVERING MOST OF THE GROUND.
AG212: TREES ABSENT OR WIDELY SCATTERED, GROUND COVERED BY LOW LYING VEGETATION.
AG222: CROWNS OF INDIVIDUAL PLANTS DISCERNIBLE, TEXTURE COARSE AND MOTTLED.

Figure 6: A Decision Tree Representation of the Dichotomous Photo Interpretation
Key for Crop Classification (Source, E.G. Mtalo, 1990, page 188).

Figure 7: Basic Operators of the LEARN Subsystem.

A typical query is processed as illustrated below:
Choose Option: S
Enter Subject: BAND 1

BAND 1 has spectral wavelength 0.45 to 0.52
BAND 1 has blue nominal spectral location.
BAND 1 application is sensing in the chlorophyll absorption region.
BAND 1 has spectral wavelength 0.45 to 0.52 or blue nominal spectral location.

Complex queries involving class relationships and limited inheritance of class attributes by class members can also be processed by the LEARN sub-system. During knowledge compilation multiple assignments of attributes (on the same relationship) to the same object are generalized into a single fact, that is the statement, "plot A56 has area=50 ha"; "plot A56 has average slope=15 degrees"; "plot A56 has curve number=80"; are replaced by "plot A56 has area=50 ha or average slope=15 degrees or curve number=80". Conversely, occurrences of multiple facts bearing the same relationship to a single attribute are replaced by a single expression in which the "subject" consists of individual subjects concatenated by the "or" operator.

3.3 Processing and Manipulation of Fuzzy Knowledge.

It is an established fact that human beings often express knowledge in fuzzy terms (Zadeh, 1989). In certain fields of application experts rely to a large extent on fuzzy, non-precise facts and information during the solution of complex problems. This is also true in the soil erosion domain where crucial inputs, such as, soil characteristics, plant cover, topography, erosion state etc. are often expressed in fuzzy non-precise terms.

Processing of fuzzy knowledge by computer systems poses three important problems, specifically, the representation and storage of non-precise (fuzzy) information and data, representation of the uncertainty inherent in fuzzy information and data, and retrieval and analysis of fuzzy information and data. Conventional databases cannot store or query fuzzy information since they impose a strict format for data entry and query. Partial solution to the problem of fuzzy queries has however been achieved by the extension of the relational data model as in Kandel (1986). Although such solutions enable processing of fuzzy queries they still do not permit the storage of fuzzy data such as for example "terrain slope=slightly less than 15 degrees".
The fuzzy knowledge processing operator is based on a new method for processing fuzzy expressions based on the concept of fuzzy geometric partition of the search space (Mtalo, 1990). During a query session the FUZZ module parses and tests both the query and examined database objects against a limited set of fuzzy expressions. If no fuzzy expression is found, the module passes the query to the normal query processor, otherwise, it performs a fuzzy object comparison in order to locate the matching database object.

Using this mechanism it is possible to store and query non-precise information provided by soil erosion domain experts without the loss of information associated with attempts to translate fuzzy expressions into exact or precise facts.

4. CONCLUDING REMARKS.

This paper has explored the utility of the knowledge based systems approach in the solution of soil erosion problems. The paper discussed briefly data requirements and information processing issues relevant to the introduction of the technology in soil loss estimation and modelling. A feasible method for the representation and manipulation of fuzzy information in the soil erosion domain was also developed.

An experimental knowledge based system prototype was also developed from basic principles and its application demonstrated in the soil loss estimation and modelling applications. The system, consisting of a unique combination of four easily accessible information management tools, demonstrates the viability of the knowledge based approach in general.

Although there has not been much progress in the development and use of expert systems in the soil erosion domain, the complexity of the problem beggars the adoption of knowledge based systems in this area. Also, because of the broad nature of the soil erosion problem, a multi-disciplinary approach to the development of soil erosion expert systems is strongly recommended.

In conclusion SLEMS four sub-systems offer easily manageable functions for solving simple soil erosion related problems. Its ability to manipulate vague information provides a partial solution to the problem of handling fuzzy data and fuzzy queries. In addition its is strongly argued that knowledge based systems are a useful vehicle for inter-disciplinary technology transfer. Finally, on the basis of the response from a multi-disciplinary group of experts from Tanzania, the use of the expert systems technology in developing countries is not only viable but desirable.

REFERENCES.


