Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar

Spanish Journal of Agricultural Research 2008 6(2), 252-263
ISSN: 1695-971-X

Building a fuzzy logic information network and a decision-support system for olive cultivation in Andalusia

G. Delgado[1], V. Aranda[2]*, J. Calero[2], M. Sánchez-Marañón[1], J. M. Serrano[3], D. Sánchez[4] and M. A. Vila[4]

[1] *Department of Pedology and Agricultural Chemistry. University of Granada. Spain*
[2] *Department of Geology. University of Jaén. Spain*
[3] *Department of Computer Science. University of Jaén. Spain*
[4] *Department of Computer Science and AI. University of Granada. Spain*

Abstract

In Southern Spain, olive (*Olea europaea* L.) growing is an important part of the economy, especially in the provinces of Jaén, Córdoba and Granada. This work proposes the first stages of an Information and Decision-Support System (IDSS) for providing different types of users (farmers, agricultural engineers, public services, etc.) with information on olive growing and the environment, and also assisting in decision-making. The main purposes of the project reported in this paper are to process uncertain or imprecise data, such as those concerning the environment or crops, and combine user data with other scientific-experimental data. The possibility of storing agricultural and ecological information in fuzzy relational databases, vital to the development of an IDSS is described. The information will be processed using knowledge extraction tools (fuzzy data-mining) that will allow rules on expert knowledge for assessing suitability of land to be developed and making thematic maps with the aid of Geographic Information Systems. Flexible querying will allow the users to collect information interactively from databases, while user information is constantly added. Flexible querying of databases, land suitability and thematic maps may be used to help in decision-making.

Additional key words: expert systems, flexible querying, fuzzy data-mining, fuzzy relational databases, knowledge extraction, user knowledge.

Resumen

Construcción de un sistema de información y de ayuda a la decisión mediante lógica difusa para el cultivo del olivar en Andalucía

El cultivo del olivo (*Olea europaea* L.) tiene una enorme importancia económica en la zona sur de España y concretamente en las provincias de Jaén, Córdoba y Granada. En este trabajo se propone la construcción de un sistema de información y ayuda a la toma de decisión (IDSS) que permita en el futuro a distintos tipos de usuarios (agricultores, agrónomos, administraciones públicas, etc.) obtener y manejar información sobre el cultivo de olivar y el soporte ambiental del mismo, así como ayudar en la toma de decisiones. Los principales objetivos desarrollados en este trabajo son el tratamiento de datos inciertos e imprecisos, como es el caso de la información ambiental y sobre cultivos, y la fusión de datos sobre cultivo y otros de carácter científico-experimental. Se describe la posibilidad de almacenar la información de carácter agronómico y ecológico en bases de datos relacionales, que es vital para el desarrollo de un IDSS. La información será procesada a través de herramientas de extracción de conocimiento (minería de datos difusa) y permitirá sobre la base del conocimiento experto el desarrollo de reglas para la clasificación de aptitud del terreno y para la obtención de mapas temáticos con la ayuda de Sistemas de Información Geográfica. La consulta flexible permitirá a los distintos usuarios la consulta interactiva de toda la información almacenada en las bases de datos, así como una implementación constante de las mismas. La consulta flexible de bases de datos, la idoneidad de los terrenos y los mapas temáticos pueden ser de gran utilidad en la toma de decisiones.

Palabras clave adicionales: bases de datos relacionales difusas, conocimiento de usuario, consulta flexible, extracción de conocimiento, minería de datos difusa, sistemas expertos.

* Corresponding author: varanda@ujaen.es
Received: 05-02-07; Accepted: 26-02-08.

Introduction[1]

Decision-support in agriculture affects both crops in all their phases (planting, management, costs, etc.) and distribution and organization of production systems. Decision-Support Systems (DSS), as expert systems, were originally developed to facilitate the application of crop models in a system approach to agriculture research. They were also motivated by a need to integrate knowledge about soil, climate, crops and management for making better decisions about transferring production technology from one location to others where soil and climate differed (Jones *et al*., 2003). The DSS approach provides a framework to research and understand how the system and its components function. This knowledge is then fed into models that predict the system's behavior for given conditions (Jones *et al*., 2003). Land evaluation based on physical and socioeconomic data is therefore a vital tool for decision-making, since it assesses the suitability of different agro-ecological systems in an area by analyzing the relationships between the variables affecting these systems (Sys *et al*., 1991).

One design task required for a fully operative DSS is the creation and management of databases able to integrate different information and knowledge resources in a decision-making context (Harrison, 1991; Marakas, 1999; Haag *et al*., 2000; De la Rosa *et al*., 2004). In this sense, it is very recommendable to involve users in the development and implementation of DSS tools (Matthies *et al*., 2007). User (such as the farmer) and technician or researcher's knowledge can be used as a basis for modeling tasks and integrating models created while designing a DSS.

However, to infer relevant knowledge from this information, knowledge discovery (knowledge extraction) with data-mining techniques (extraction rules), which have been demonstrated to be the best tools for agricultural and environmental systems, are used (Hoshi *et al*., 2000; Poch *et al*., 2004; Kawano *et al*., 2005).

Furthermore, in an agro-environmental context, the information derived from the vagueness of human reasoning, while experiencing and interpreting the complexity of the real world is uncertain and imprecise (Bragato, 2004). This is especially critical in the case of the farmer's knowledge of his/her agricultural system.

Fuzzy logic provides an effective decision-making tool for dealing with this kind of imprecision and uncertainty (Groenemans *et al*., 1997; McBratney and Odeh, 1997; Center and Verma, 1998; Kuo and Xue, 1998; Evsukoff *et al*., 2000; Geneste *et al*., 2003). First proposed by Zadeh in his Theory of Fuzzy Subsets (Zadeh, 1965), fuzzy logic works with rough data and linguistic values of variable and imprecise relationships to find the best solution or alternatives to complex problems (Panigrahi, 1998), such as those in agricultural systems. Fuzzy logic has been used successfully as a prediction tool in agriculture, for example, in sugar production (Petridis and Kaburlasos, 2003) and livestock raising (Lacroix *et al*., 1998).

The use of fuzzy data-mining could increase the flexibility and adaptability of rule-based DSS, and inference from fuzzy sets could be an alternative to current methods (De la Torre *et al*., 2005).

The purpose of this study is to establish the basis for creation of a future rule-oriented, integrated Information and Decision-Support System (IDSS) for olive (*Olea europaea* L.) crop management in the region of Andalusia, based on optimum deployment of resources according to their potential, e.g., sustainable land management systems. The paper summarizes the information collected (highlighting its imprecise nature), its representation in fuzzy relational databases, how flexible queries are made, different kinds of knowledge are merged and knowledge is extracted (using fuzzy data-mining). It describes the main features of this particular implementation, and establishes the basis for the development of a future complete IDSS.

Premises for the development of the IDSS

For development of a suitable IDSS, it is assumed that optimum cultivation is a system with the highest productivity compatible with maintaining the suitability of the soil for cultivation. To achieve this, a number of objectives must be reached: a) maximum profit; b) products of maximum quality; c) minimum cost, for

[1] Abbreviations used: ALES (land evaluation system) AU (agropedological unit), CF (certainty factor), CGI (common gateway interface), DB (database), FAO (Food and Agricultural Organization), FMB (fuzzy meta-knowledge base), FQ (fuzzy queries), GIS (geographical information systems), IDSS (information and decision-support system), PCA (principal components analysis), UER (user evaluation rules).

which totally mechanized olive cultivation is required; d) absence of significant limitations for the crop, so that this is not in a marginal situation; e) the use of good agricultural practices, understood as being the necessary cultural maintenance without cutting profits or increasing environmental risks; f) land use which protects the soil.

The development of this type of expert system requires the operations shown in Figure 1. Data collection for current and potential systems of olive cultivation may be carried out through the bibliography or directly. Working with agronomic systems involves the use of different types of knowledge such as scientific knowledge and that derived from users. Knowledge in the form of bibliographical and direct data on cultivation systems is stored in soil and cultivation databases and in digital maps.

Expert knowledge rules are found by extracting knowledge from the information. An example of rule extraction by means of fuzzy data-mining is explained in the last section of the manuscript. These will allow establishing a knowledge tree based on fuzzy logic and use an expert evaluation system, thus creating a computerized suitability classification based on fuzzy rules.

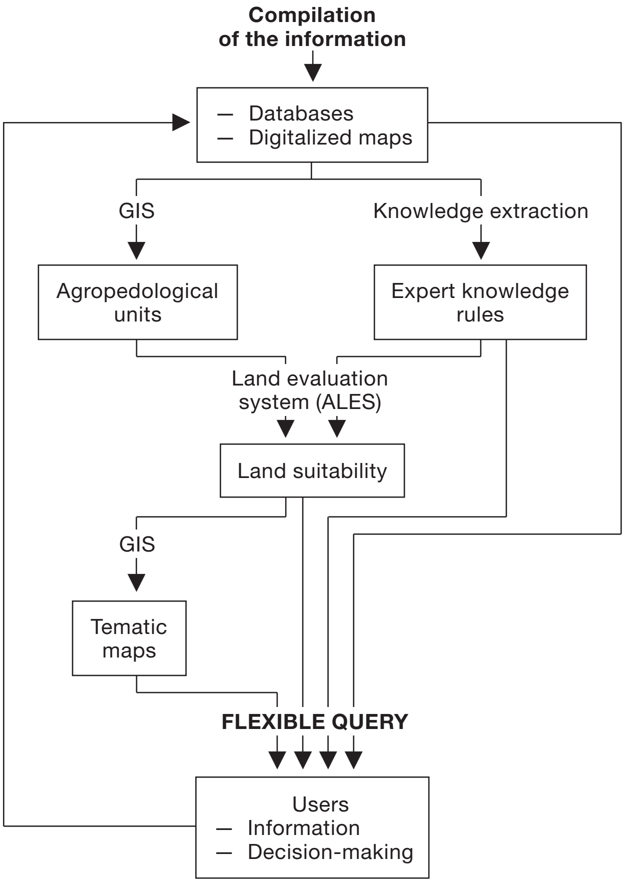


Figure 1. Operations of the information and decision-support system.

From the databases and the digitalized maps, homogeneous soil-cultivation areas called «agropedological units» (AU) are established by merging soil and cultivation knowledge. These units are assessed using suitability classification, thus allowing classification of the suitability of the current and potential systems for olive cultivation.

Using Geographical Information Systems (GIS) all the results from the assessment and everything held in databases and digital maps may be expressed spatially (thematic maps). Its usefulness in land-use planning is greatest and allows model outputs, such as constraints on agriculture and site-specific best-management practices, to be identified in a spatially explicit manner (Smith *et al*., 2000).

Description of the information

Type, source and spatial representation of the information

For an agricultural IDSS, in addition to socioeconomic information, information on soils (in the wide sense, including climate, topography, etc.) and crops is required. The former consists of pedological (soil considered in the strict sense of «individual soil»), climatic, topographical and geological data and can be found in the bibliography in the form of thematic maps, databases, digital terrain models, etc. This information is represented spatially as soil unit maps, climatic areas with the same temperature or rainfall, isocline zones, lithostratigraphic units, etc. The information on crops is of an agronomic (varieties of plants cultivated, irrigation systems, fertilizers, etc.) and socioeconomic nature. The cultivation data are generally compiled directly through interviews with farmers and indirectly from agricultural reports. Surveys function as a result of rural people in many developing countries having a rich understanding of their resources (Thrupp, 1989; Warren, 1989).

The survey used in this IDSS consisted of three sections: A) location and general data (a first section with survey control information together with geographical data); B) crop management (including attributes of land area, number and characteristics of plants, treatment and care, and production data) and C) soil data (general characteristics, such as soil depth, hardened layer depth, mean slope, soil texture, etc.).

In the survey, the spatial unit of information on the crop is the plot or farmer's property, while other data are given by district or even province (local or regional level).

Inaccuracy and uncertainty of the information

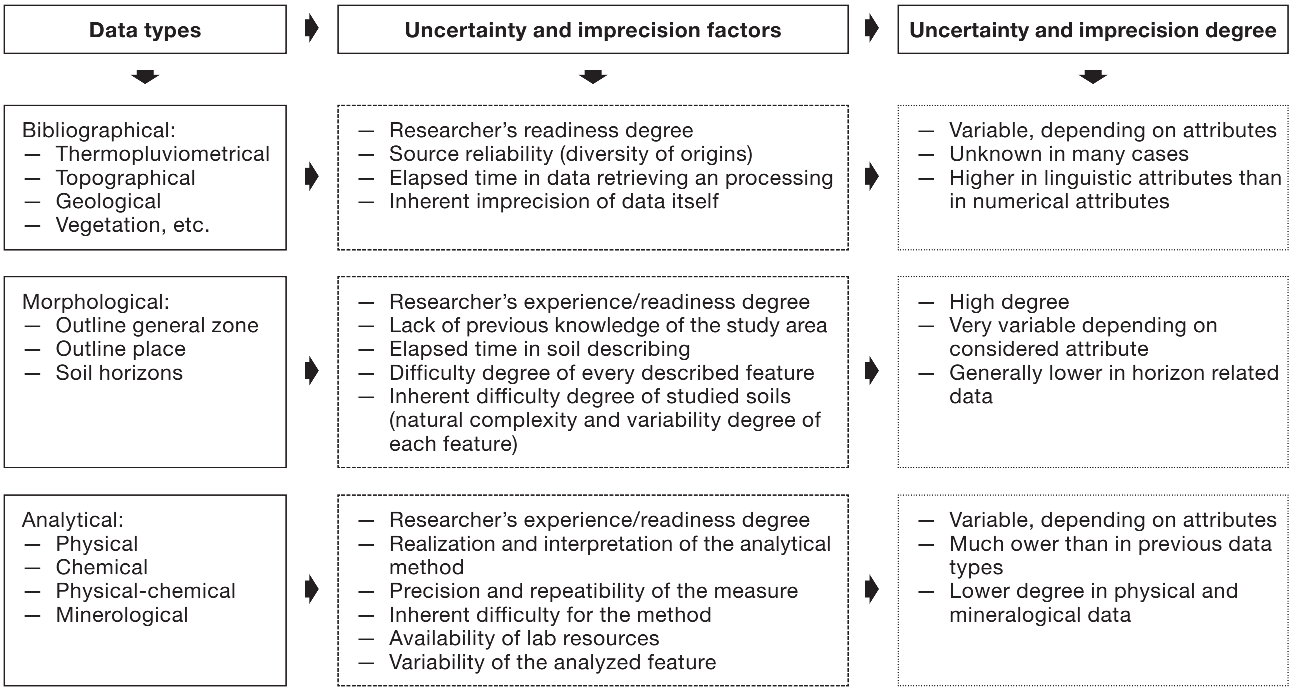
The first source of uncertainty in the farmer survey is due to the fact that only a small percentage of the farmers in the province of Granada was interviewed (210 surveys); Rodríguez *et al*. (1998) state that the minimum percentage of interviews is a function of the variability of the agricultural systems under study. Since knowledge about this variability requires surveys a vicious circle is created.

Most of the farmers interviewed have highly fragmented properties; the survey is carried out on the most representative area which may lead to a distortion in size of the property. The survey may also create uncertainty or inaccuracy as a result of linguistic problems since some terms are not common for the whole of the area surveyed. The cost of many agricultural activities is difficult to calculate, especially when these tasks are carried out by the farmer and his family. Farmers are also sometimes uncertain about the prices of fertilizers and other products. Another problem is the tendency, for socio-cultural reasons, neither to disclose the real agricultural practices carried out, nor their cost. It should always be remembered that the farmer will have a different concept about the soil to that held by the soil scientist.

Soil databases store two essential types of attributes regarding their obtention: bibliographical and experimental, and, within the latter, morphological and analytical one (soil horizons, physical, chemical, etc). In Figure 2 the types of data are related to the factors and their uncertainty and inaccuracy. It should also be noted that much of the information regarding soils is qualitative and is thus exposed to the researcher's subjectivity and, furthermore, its mathematical processing is difficult (Webster, 1977; Webster and Oliver, 1990). The analysis of an object, soil, whose variability is produced in a «*continuum*», is difficult, since it requires the use of new methods based on fuzzy logic and fuzzy groups (McBratney and De Gruijter, 1992).

Imprecise information representation using fuzzy logic

One of the most widespread methodologies in information organization and processing is the rela-



Data types	Uncertainty and imprecision factors	Uncertainty and imprecision degree
Bibliographical: — Thermopluviometrical — Topographical — Geological — Vegetation, etc.	— Researcher's readiness degree — Source reliability (diversity of origins) — Elapsed time in data retrieving an processing — Inherent imprecision of data itself	— Variable, depending on attributes — Unknown in many cases — Higher in linguistic attributes than in numerical attributes
Morphological: — Outline general zone — Outline place — Soil horizons	— Researcher's experience/readiness degree — Lack of previous knowledge of the study area — Elapsed time in soil describing — Difficulty degree of every described feature — Inherent difficulty degree of studied soils (natural complexity and variability degree of each feature)	— High degree — Very variable depending on considered attribute — Generally lower in horizon related data
Analytical: — Physical — Chemical — Physical-chemical — Minerological	— Researcher's experience/readiness degree — Realization and interpretation of the analytical method — Precision and repeatibility of the measure — Inherent difficulty for the method — Availability of lab resources — Variability of the analyzed feature	— Variable, depending on attributes — Much ower than in previous data types — Lower degree in physical and mineralogical data

Figure 2. Imprecision and uncertainty factors in soil data.

tional database model. Data is handled in registers, which are grouped in tables or relations with a defined operations set in order to obtain, reorganize and/or alter information stored in the database.

Originally, the relational model did not allow imprecise or uncertain information to be handled. However, in recent years, researchers have tackled the problem of how to make this model more flexible in order to allow this feature. Fuzzy logic is thought to be able to deal with this kind of problem both effectively and efficiently.

Over the last two decades various improvements have been made to the relational model so that it may manage imprecision and uncertainty using fuzzy logic. Different models have been proposed. The GEFRED model (Medina *et al.*, 1994) is generally the most well-known as it handles imprecision and uncertainty simultaneously. It also proposes a special approach in the query process to the Fuzzy Relational Database Management System (FRDBMS), in which any information request to the database is reduced to a query in relational terms. The model considers nearness relations over attribute domains and also the possibility of setting different fulfillment degrees for any of the attributes in the query.

Fusion of the information on crops and soils. Spatial representation of the data (GIS)

The fusion of data on a spatial level was resolved by establishing AUs for the province of Granada. The AU is an abstract concept defined by soil typologies and other physical parameters (slope and temperature) which have a decisive influence on olive cultivation. The AUs were established by superimposing three different maps: soils, mean annual temperature and slope. The aim of this division of the land is to establish areas with a suitable size in order to reduce the degree of variability of the conditions for this crop and to ensure that the number of data for each kind of unit is sufficient. The databases for the soils and the interviews were spatially associated within the AUs.

The soil map used was that for the province of Granada on the 1:200,000 map scale (Pérez-Pujalte, 1980). This map is based on geology and physiography, which results in the grouping of soil typologies which have different classifications but with great similarities in terms of morphology and properties (soil texture, organic carbon content, pH, etc.).

The temperature map was established using the temperature/altitude correlation, with this province having very high correlation coefficients. According to these correlations, the limits for olive cultivation were –7°C for the mean minimum absolute temperature of the coldest month and 13°C for the annual mean, giving the following thresholds: higher than 1,250 m, low suitability; from 1,250 down to 800 m, reasonable suitability; from 800 down to 275 m, optimum suitability, and, lower than 275 m, reasonable suitability.

Slope conditions affect certain aspects of cultivation such as: soil management (mainly irrigation and mechanization) and risks of degradation (mainly hydric erosion). Using the limits established by the Food and Agricultural Organization (FAO) for the slope, four classes of limitation can be established, these being the four map units of slope on the map: less than 6% slope, no limitation; from 6 to 13%, moderate limitations; from 13 to 25%, severe limitations, and, more than 25%, very severe limitations.

The AU can be considered as a «unit of information», that is, it cannot be defined by itself but by its attributes. It is not intended to be a «natural» unit of the landscape but rather a division which mixes the soil map unit, based on a natural soil classification (FAO, 1974), with artificial units for the landscape established using utilitarian criteria (risk of frost, latent period, possible mechanization, etc.). However, since these utilitarian criteria are expressed as thresholds of altitude or slope, natural variability is also involved. Thus, the AU is a division of the landscape which is essentially natural; it is not exactly a «unit of land type» as defined by FAO (1974).

Figure 3 shows how the AU is interpreted graphically. Once the AU is established, the agronomic information (surveys of farmers) and information on soils and other databases are incorporated. The AU can be reinterpreted from the point of view of olive cultivation (including utilitarian information) with the aid of the suitability classification. The suitability of these AUs can be used for decision-making, geographically speaking, using the systems for flexible querying.

The incorporation of GIS as tools for modeling and spatial analysis is vital for any project where georeferenced information must be integrated (Longley *et al.*, 1999). With a view to handling the mapping information and establish the AUs, ARC/INFO v.7.02 for Windows NT was used owing to its capacity for the integration of database systems external to the system,

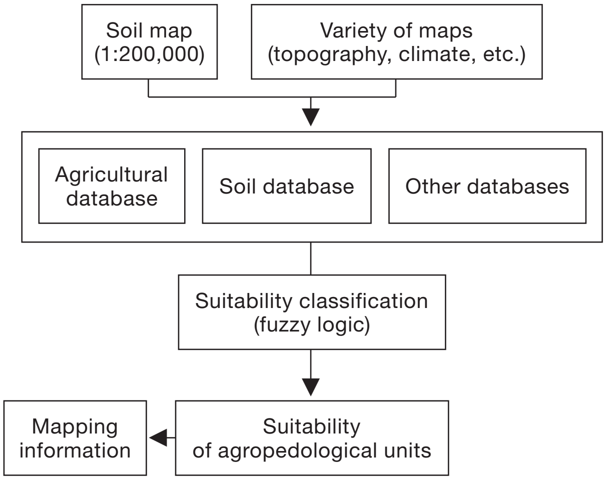


Figure 3. Use of agropedological units.

such as Oracle, and with special software for the construction of models of land assessment, such as the ALES system. The initial mapping information was incorporated into the system by digitalizing the corresponding maps or was acquired directly in digital formats compatible with ARC/INFO. In some cases this information was modeled directly as vectorial covers, for example, the maps for soil and soil uses. In other cases raster models in a grid or lattice format of ARC/INFO were used, as is the case of the digital terrain model with a pixel size of 20 m. The climatic data (mean annual temperature and mean annual precipitation) were initially modeled on a digital terrain model using linear correlation equations obtained from 40 meteorological stations in or around the province of Granada. Different raster models were constructed for the different climatic parameters modeled. Vectorial covers were generated from this model that corresponds with the limits of the climatic parameters employed in the mapping.

Slope maps in the form of vectorial covers were also generated from the digital terrain map, using the slope margins listed previously.

The databases associated with the vectorial covers were constructed and handled in Oracle, with the covers being connected through the corresponding feature attribute tables.

Once all the information had been integrated into the GIS, spatial analysis tools such as topological analysis (spatial join, proximities, buffering, etc.) of vectorial covers together with logical operations on the feature attribute tables were used to obtain maps of agropedological units, thematic maps, etc.

Flexible querying and knowledge extraction

Operations on imprecise data

Many different criteria have been considered in order to decide whether to model information as precise or imprecise. On the one hand, there are numerical values. When these are obtained from a reliable source, such as an analytical study, it is not necessary to store them as fuzzy data, for instance, % of nitrogen in soil. On the other hand, there are no exact measurements for the products used in fertilizing or fumigation and the quantities are based on the farmer's experience. This is an additional source of imprecision, and data must be handled and stored appropriately. Information on other attributes such as land slope is supplied using the linguistic labels (flat, moderately sloping, sloping) which are closer to natural language, although the underlying domain may be numerical.

These can therefore be modeled as fuzzy values, using trapezoidal or interval functions as possibility distributions. Even when the attributes are handled as labels, a query can be made regarding the numerical (fuzzy) values. In order to do so, information about the match between linguistic labels and their associated value sets must be stored in the FRDBMS Catalog, more precisely, in the Fuzzy Meta-Knowledge Base (FMB).

There are also other types of imprecise attributes, which are those represented as a set of scalar non-numerical values by means of a nearness relationship. For example, the attribute «land tractorability» has a class set, defined as (high, medium, low), following an intuitive order relationship. The nearness matrix between these categories is stored in the FMB.

In the soil database, most of the attributes are defined on a categorized domain due to the standards proposed by the FAO. As existing databases are modeled in this way it is impossible for imprecision or uncertainty to be handled. However, it is possible to handle these features. Although stored data is «crisp», uncertainty can be added when flexible queries are performed, thereby improving the model so that information may be obtained.

Thus, it may be said that two types of fuzzy data processing are performed. One of these is called *a priori* handling, in which there is fuzzy data since this is how it was stored in the DB. The other is called *a posteriori* handling and takes place in the query process. Here, data that have really been stored as precise can be fuzzified.

The flexible query

A typical system for handling bases provides procedures for the storage, access and modification of the information. Through the facility of access to the information the user can obtain information contained in the database (DB). It is even possible to specify the criteria with which the information is selected; or to demand that only the information which fulfills a certain condition (degree of fulfillment) is returned to the user.

Furthermore, when part of the information is fuzzy it is necessary to provide methods which give good access to this information. One way to achieve this is to flexibilize the query. This gives the language for querying the DB the ability to handle imprecise information and to express fuzzy data in natural language terms.

For example, a typical query addressed to the DB: «Tell me which farms have an annual production of more than 3000 kg», would read, «*SELECT farm_name, annual_production FROM farm WHERE annual_production>= 3000*».

Supposing that it is needed to know which farms have a «high» annual production, the flexible query would be as follows: «*SELECT farm_name, annual_production FROM farm WHERE annual_production = HIGH*».

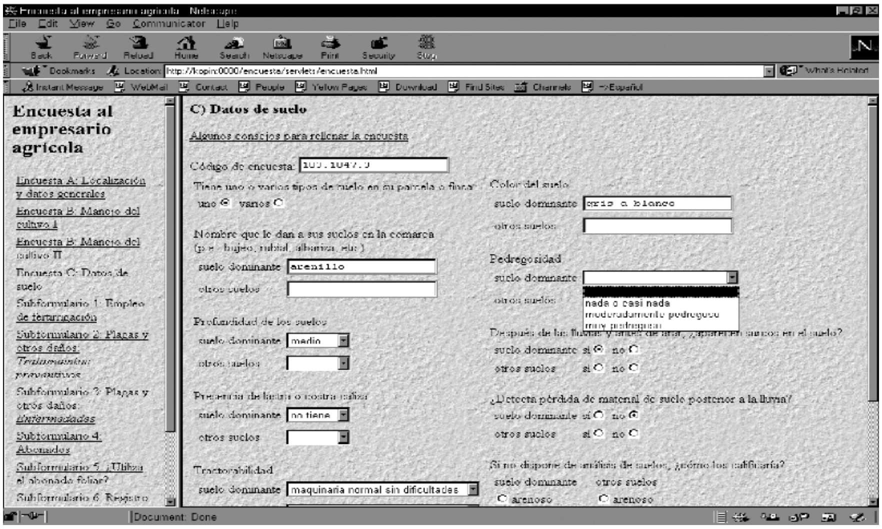
The common term «high» is extremely imprecise. How many olives should be obtained for production to be considered «high»? Internally, the DB system could model this characteristic in different ways, although this processing should be hidden from the user, who is only interested in finding out which farms have high production. It should also be noted that this flexible query can also be carried out both on attributes of a fuzzy nature and on those with a perfectly defined value.

Different models of DB have considered the problem of flexible query, one of which, the GEFRED model (Medina *et al*., 1994), mentioned previously, is the most widely known one.

In the process of implementation of the information, the modules for data input and flexible query should be examined.

Data input module

Using this first module, users can easily enter new data. It consists of two clearly different parts. The visible part is a set of HTML forms (Fig. 4), each associated to a table in the database. Users can access them via internet using any web browser. Under this, there is a Java-based application, comprising a servlet (current replacements for CGI, an usual technology employed in web servers) and an insertor class. This part processes the sentences and communicates with the FRDBMS.


Encuesta al empresario agricola - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://kopin:0000/encuesta/servlets/encuesta.html
What's Related
Instant Message WebMail Contact People Yellow Pages Download Find Sites Channels ->Español
Document: Done

Encuesta al empresario agrícola

Encuesta A: Localización y datos generales
Encuesta B: Manejo del cultivo I
Encuesta B: Manejo del cultivo II
Encuesta C: Datos de suelo
Subformulario 1: Empleo de fertirrigación
Subformulario 2: Plagas y otros daños: *Tratamientos preventivos*
Subformulario 3: Plagas y otros daños: *Enfermedades*
Subformulario 4: Abonados
Subformulario 5: ¿Utiliza el abonado foliar?
Subformulario 6: Registro

C) Datos de suelo

Algunos consejos para rellenar la encuesta

Código de encuesta: 100.1047.0

Tiene uno o varios tipos de suelo en su parcela o finca:
uno ● varios ○

Nombre que le dan a sus suelos en la comarca (p.e. bujeo, rubial, albariza, etc.)
suelo dominante: arenillo
otros suelos:

Profundidad de los suelos
suelo dominante: medio
otros suelos:

Presencia de lastra o costra caliza
suelo dominante: no tiene
otros suelos:

Tractorabilidad
suelo dominante: maquinaria normal sin dificultades

Color del suelo
suelo dominante: gris a blanco
otros suelos:

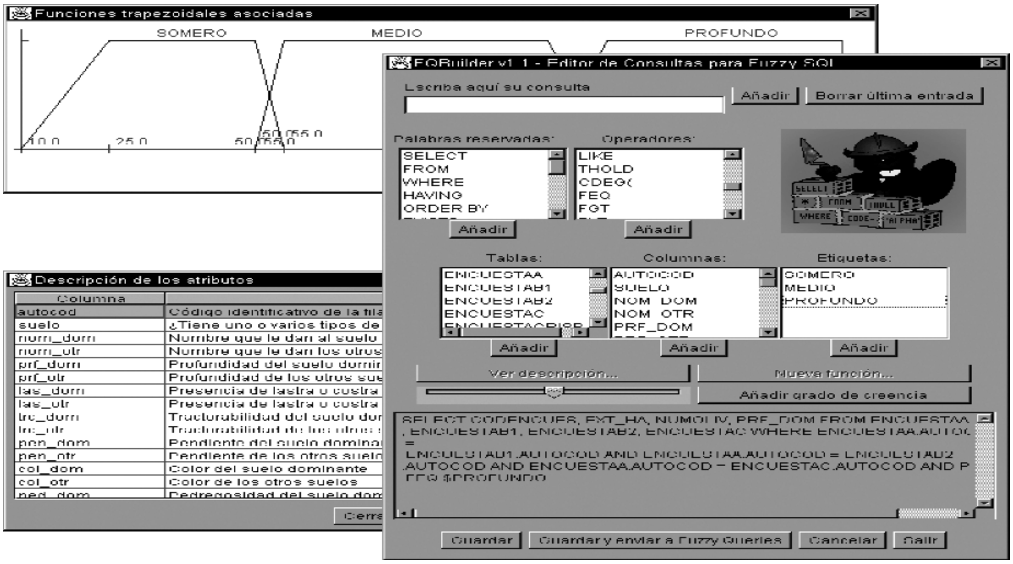
Pedregosidad
suelo dominante:
otros suelos:
nada o casi nada
moderadamente pedregoso
muy pedregoso

Después de las lluvias y antes de arar, ¿aparecen surcos en el suelo?
suelo dominante sí ● no ○
otros suelos sí ○ no ○

¿Detecta pérdida de material de suelo posterior a la lluvia?
suelo dominante sí ○ no ●
otros suelos sí ○ no ○

Si no dispone de análisis de suelos, ¿cómo los calificaría?
suelo dominante otros suelos
○ arenoso ○ arenoso

Figure 4. HTML form model.



Columna	
autocod	Código identificativo de la fila
suelo	¿Tiene uno o varios tipos de
nom_dom	Nombre que le dan al suelo
nom_otr	Nombre que le dan los otros
prf_dom	Profundidad del suelo domin
prf_otr	Profundidad de los otros sue
las_dom	Presencia de lastra o costra
las_otr	Presencia de lastra o costra
trc_dom	Tracturabilidad del suelo dom
trc_otr	Tracturabilidad de los otros s
pen_dom	Pendiente del suelo dominan
pen_otr	Pendiente de los otros suelo
col_dom	Color del suelo dominante
col_otr	Color de los otros suelos
ped_dom	Pedregosidad del suelo dom

```
SELECT CODENCUES, EXT_HA, NUMOLIV, PRF_DOM FROM ENCUESTAA
, ENCUESTAB1, ENCUESTAB2, ENCUESTAC WHERE ENCUESTAA.AUTOC
=
ENCUESTAB1.AUTOCOD AND ENCUESTAA.AUTOCOD = ENCUESTAB2
.AUTOCOD AND ENCUESTAA.AUTOCOD = ENCUESTAC.AUTOCOD AND P
FEQ $PROFUNDO
```

Figure 5. Construction of a query with FQBuilder.

The servlet application manages the web pages, takes data from forms and informs the users about the success or failure of the transactions performed. In order to accomplish these tasks it transmits the content of the forms to the insertor class, and receives and manages the success or error messages sent by the server.

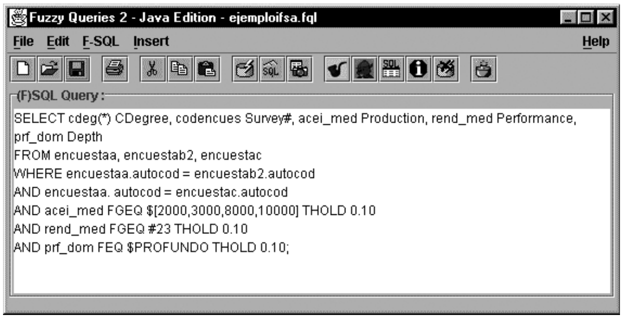
Data is processed within the insertor class in order to build the insertion sentence. Communication with the FRDBMS is also maintained here, and any sentences sent and any resulting messages received.

The module is sufficiently general to be adapted to any given set of tables, which may or may not contain fuzzy values. For this case, two versions have been implemented. The first is for the survey database and the second for the soil database. Another added feature is that of allowing simple queries and on-line updates of previously stored data using the same interface. If more complex queries are to be performed, e.g. flexible queries, another tool must be used, as explained below.

Flexible query module

This module is responsible for displaying information, which has previously been stored in the database, to the user according to a flexible (or non-flexible) query specified by the user. Although two different applications were originally used in this module, these were later integrated into one. The original program, FQ (Fuzzy Queries) (Galindo *et al.*, 1998), was improved with the name of FQBuilder (Fig. 5). This was programmed in Visual Basic. Firstly, a separate Java-based application was developed in order to improve FQ. Following a short period of experimentation, a more complete and powerful application was developed. FuzzyQueries2 (Fig. 6) takes the best of the previous releases and adds new features. Using this application, users can easily edit and build flexible queries in FSQL.

FuzzyQueries2 can display a set of reserved words, common operators and objects in the database (tables, columns and linguistic label sets), which users can take advantage of. Using FuzzyQueries2 users can query the system catalog, and, within this, the Fuzzy Meta-Knowledge Base (FMB). In particular, FuzzyQueries2 displays the existing matches between linguistic labels and trapezoidal distributions, both defined on fuzzy attributes in the databases. Users can define their own possibilistic distributions using the graphic interface.



```
SELECT cdeg(*) CDegree, codencues Survey#, acei_med Production, rend_med Performance,
prf_dom Depth
FROM encuestaa, encuestab2, encuestac
WHERE encuestaa.autocod = encuestab2.autocod
AND encuestaa. autocod = encuestac.autocod
AND acei_med FGEQ $[2000,3000,8000,10000] THOLD 0.10
AND rend_med FGEQ #23 THOLD 0.10
AND prf_dom FEQ $PROFUNDO THOLD 0.10;
```

Figure 6. FuzzyQueries2 main window.

Once the sentence has been built, it is sent to the FRDBMS, where it is analyzed and any syntax errors are found, before being executed. The results are then displayed to users.

Knowledge extraction

Using the query (flexible or not) on the database being used it is possible to obtain extrinsic or direct information immediately. However, one of the main reasons for compiling data on olive cultivation and soils in Granada is to obtain intrinsic or hidden information and to find relationships which are not evident between attributes which are not known in advance. To obtain this new class of information, knowledge extraction and data-mining techniques are used (Cubero *et al*., 1995; Carrasco *et al*., 2000; Delgado *et al*., 2000).

The tools for knowledge extraction are extremely powerful but must be refined beforehand in order to orientate the search without increasing excessively experimental time while extracting genuinely useful information for the user. It is thus necessary to carry out a first exploratory analysis of the data to select the attributes which best define the information. A variety of useful statistical tools are available for this purpose.

The first step is to carry out a multivariate analysis by means of a principal components analysis (PCA) which restricts the initial group of attributes to a subgroup which is easier to handle. Once these principal components are known, providing initial information on the relevance of the attributes, the elements can be grouped into clusters or classes according to these principal components (Calero *et al*., 2005). These groups are an additional source of information, in this case for the similarities between elements (Delgado *et al*., 1999).

From this first exploratory analysis of the data were obtained a restricted group of relevant attributes and the clusters in which the elements were divided. A knowledge extraction technique can now be used. One option is to make use of the classes obtained in the second step to use an algorithm based on classification trees, such as C4.5 or ID3 (Carrasco *et al*., 1998). Another option is to search for intrinsic relationships between attributes using a data-mining tool such as rules of association (Delgado *et al*., 1999, 2000). The results obtained can then be compared with the expected results, providing a basis for further study.

Example of fuzzy data-mining application over user knowledge in olive cultivation

As described earlier, the information was obtained from the knowledge that farmers supplied (user knowledge) by means of surveys about farm management, soils and the environment. The user data obtained from a survey with 126 variables carried out on 210 olive grove farmers in the Province of Granada (southern Spain) was stored in a GEFRED model database and processed using data-mining. This information was pre-processed in order to select the most relevant attributes. The resulting data were fuzzified, by defining linguistic labels over numeric attributes [for each numeric attribute, is defined a set of linguistic labels (high, medium, low) over the numeric domain] and fuzzy relations over scalar attributes. Then, a fuzzy association rule extraction process was applied in order to obtain user evaluation rules (UER).

Using 34 variables selected following a cleaning process, 1420 fuzzy association rules were obtained in which the antecedent is a variable of management, soil, or environment, and the consequent was: olive fruit production (kg ha^{-1}), % of oil in olive fruit, or acidity of fruit on tree (olive oil quality). Only 182 of these rules were selected, either because the value of their certainty factor (CF) (assurance measure) was high, or because they were most interesting from the point of view of expert knowledge. From the 182 rules, 85 were deemed to be user evaluation rules indicating the suitability for olive cultivation according to the different attributes, and they were drawn up on the basis of user knowledge. Some of these rules were corroborated with the knowledge regarding olive cultivation found in the bibliography, others contradicted the said knowledge and others revealed relationships not previously described. This procedure would be deemed to be an evaluation method based on empirical data.

As follows, are commented some rules that might be of interest following the scheme: (Antecedent) → (Consequent) CF, comparing them with expert rules taken from the bibliography. The consistency of the method for obtaining fuzzy association rules by means of fuzzy data-mining was confirmed with the following rule: (Mean minimum temperature in the coldest month = high) → (Mean altitude = low) CF = 1. The association between these two variables has universal validity. The results are focused on rules in which the consequent is a production variable (olive fruit production in kg ha^{-1}), that is the most significant one.

The following rules, related with agronomical management, would be deemed to be highly interesting. This is the case for: (% planting frame of less than 10 m = medium) → (production = low) CF 0.60. This rule, according to the bibliography, matches the one found in Barranco *et al*. (1999). The authors consider the optimal planting frame to be between 200 and 300 trees ha^{-1}, i.e., an approximate distance between trees of 10 m or less.

Some UER are obvious, as is the case of (% of young olive trees = high) → (production = low) CF 0.70 or (% of olive trees in full production = low) → (production = low) CF 0.44.

Other UER seem to be contradictory according to other sources of knowledge. This is the case of the negative relationship between irrigation and production [(irrigation = with irrigation) → (production = low) CF 0.35], that should be positive, as described in the bibliography (Barranco *et al*., 1999), even more considering the severe summer drought in the region. This contradictory association might be explained using a third variable: the Granada olive grove system is undergoing a phase of modernization in which irrigation and the rejuvenation of the olive trees are included amongst the improvements, therefore, the olive trees with irrigation are usually the youngest ones, which are not yet in full production.

There are also UER that reveal knowledge that is difficult to interpret, although the information that they hide is interesting. For example: the rule that relates high production with a low number of olive trees [(number olive trees = low) → (production = high) CF 0.40] and its complementary rule that is the one which associates low production with a large farming area [(farming area = high) → (production = low) CF 0.49]. It should be borne in mind that the larger the farming area, the larger the number of olive trees. More research would be required to provide clearer answers and in order to validate this rule.

Some UER related with soil characteristics may also be considered very interesting. For example: (soil depth = medium) → (production = low) CF 0.45; (workability = low) → (production = low) CF 0.40; (mean slope = high) → (production = medium) CF 0.32; (stoniness = high) → (production = low) CF 0.55; (soil texture = sandy) → (production = low) CF 0.51; (erosion = with erosion) → (production = low) CF 0.30. These association rules reveal that the characteristics of the soil which influence agricultural practices, the handling of the soil or the development of the roots, are all clearly perceived by the farmer. In most cases when these soil characteristics are unfavourable (including the erosion variable), they are related to low production. The soil colour perceived and defined by the farmers also produces interesting UER. For example: (soil colour = dark grey) → (production = low) CF 0.59. This rule is difficult to explain and needs further field research.

The latter and other UER, demonstrate that the method for knowledge extraction described in this section is of an exploratory nature and so allows new hypotheses and studies to be proposed.

Conclusions

The purpose of this study was to establish the basis for a data system to assist in decision-making in the scope of olive production in the region of Andalusia (Southern Spain). To do so, intense work has been done to integrate information with different degrees of imprecision and uncertainty gathered from different experimental and empirical sources. In addition, the goal is to create an interactive system that will provide information, while simultaneously receiving information, for the different system users, by implementing a procedure for flexible querying.

An important part of the work is the design of fuzzy databases and the use of fuzzy logic for storing and processing the environmental data (mainly on soils) and agricultural data. Current database models are too strict to store imprecise and uncertain data of this kind. Moreover, fuzzy databases modeled by fuzzy logic allow flexible querying, which is an essential tool in the creation of an IDSS.

The system which it is intended to create is still in the development stage. Several stages have been completed and others are still underway. Compilation of information and database implementation is already finished. An exploratory analysis of the data has been carried out and Agropedological Units have been established using GIS. Expert knowledge rules based on the farmer survey have been extracted using fuzzy data-mining techniques (fuzzy association rules and fuzzy approximate dependencies).

Some problems have come up in the user information compilation stage and in flexible iterative querying by those users. Specifically, the number of items in the farmer survey was excessive and reiterative, for example, plot size and number of olive trees. Some of the questions asked the farmer e.g., about soil colour, are difficult to interpret as well. The query process also baffles some of the users, who lack training, since they are usually elderly and have small farms. So to solve these problems, a new grower survey is now being designed with fewer items, the questions are being rewritten, graphs and photographs are included and a more user-friendly interface is being implemented for flexible querying (both for data entry on and retrieval from the web page through an extensive client application).

Acknowledgements

This work is part of the research projects 1FD97-0244-CO3-2 (financed with FEDER funds) and CGL2004-02282BTE (Spanish Ministry of Education and Science).

References

BARRANCO D., FERNÁNDEZ-ESCOBAR D., RALLO L., 1999. El cultivo del olivo. Coedition Consejería de Agricultura y Pesca (Junta de Andalucía) and Ediciones Mundi-Prensa, Madrid, Spain. [In Spanish].

BRAGATO G., 2004. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. Geoderma 118, 1-16.

CALERO J., SERRANO J.M., ARANDA V., SÁNCHEZ D., VILA M.A., DELGADO G., 2005. Analysis and characterization of olive tree cultivation system in Granada province (South of Spain) with optimal scaling and multivariate techniques. Agrochimica XLIX, 118-131.

CARRASCO R., GALINDO J., MEDINA J.M., ARANDA M.C., VILA M.A., 1998. Classification in databases using a fuzzy querying language. International Conference of Management of Data, COMAD'98, Hyderabad, India.

CARRASCO R., VILA M.A., GALINDO J., CUBERO J.C., 2000. FSQL, a tool to obtain fuzzy dependencies. IPMU'2000, Madrid, Spain.

CENTER B., VERMA B.P., 1998. Fuzzy logic for biological and agricultural systems. Artif Intell Rev 12, 213-225.

CUBERO J.C., VILA M.A., MEDINA J.M., PONS O., 1995. Rules discovery in fuzzy relational databases. ISUMA-NAFIPS'95, Computer Soc Press, Maryland, USA.

DE LA ROSA D., MAYOL F., DÍAZ-PEREIRA E., FERNÁNDEZ M., DE LA ROSA D., 2004. A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection. With special reference to the Mediterranean region. Environ Modell Softw 19, 929-942.

DE LA TORRE M.L., GRANDE J.A., AROBA J., ANDUJAR J.M., 2005. Optimization of fertirrigation efficiency in strawberry crops by application of fuzzy logic techniques. J Environ Monit 7, 1085-1092.

DELGADO M., GÓMEZ-SKARMETA A.F., VILA M.A., 1999. Pattern recognition with evidential knowledge. Int J Intell Syst 14, 145-164.

DELGADO M., SÁNCHEZ D., MARTÍN-BAUTISTA M.J., VILA M.A., 2000. Mining association rules with improved semantics in medical databases. Artif Intell Med 587, 1-5.

EVSUKOFF A., GENTIL S., MONTMAIN J., 2000. Fuzzy reasoning in co-operative supervision systems. Control Eng Pract 8, 389-407.

FAO, 1974. Soil map of the world. FAO-Unesco, Legend, Paris.

GALINDO J., MEDINA J.M., PONS O., CUBERO J.C., 1998. A server for fuzzy SQL queries. In: Flexible query answering systems (Andreasen T., Christiansen H., Larsen H.L., eds). Lecture Notes in Artificial Intelligence (LNAI) 1495, pp. 164-174. Springer.

GENESTE L., GRABOT B., LETOUZEY A., 2003. Scheduling uncertain orders in the customers-subcontractor context. Eur J Oper Res 147, 297-311.

GROENEMANS R., VAN RANST E., KERRE E., 1997. Fuzzy relational calculus in land evaluation. Geoderma 77, 283-298.

HAAG S., CUMMINGS M., MCCUBBREY D.J., PINSONNEAULT A., DONOVAN R., 2000. Management information systems: for the information age. McGraw-Hill Ryerson Limited: Denver, USA. pp. 136-140.

HARRISON S.R., 1991. Validation of agricultural expert systems. Agr Syst 35, 265-285.

HOSHI T., SASAKI T., TSUTSUI H., WATANABE T., TAGAWA F., 2000. A daily harvest prediction model of cherry tomatoes by mining from past averaging data and using topological case-based modelling. Comput Electron Agric 29, 149-160.

JONES J.W., HOOGENBOOM G., PORTER C.H., BOOTE K.J., BATCHELOR W.D, HUNT L.A., WILKENS P.W., SINGH U., GIJSMAN A.J., RITCHIE J.T., 2003. The DSSAT cropping system model. Eur J Agron 18, 235-265.

KAWANO S., HUYNH V.H., RYOKE M., NAKAMORI Y., 2005. A context-dependent knowledge model for evaluation of regional environment. Environ Modell Softw 20, 343-352.

KUO R.J., XUE K.C., 1998. A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights. Decis Support Syst 24, 105-126.

LACROIX R., STRASSER M., KOK R., WADE K.M., 1998. Performance analysis of a fuzzy decision-support system for culling of dairy cows. Can Agr Eng 40, 139-152.

LONGLEY P.A., GOODCHILD M.F., MAGUIRE D.J., RHIND D.W., 1999. Geographical information systems. J Wiley & Sons, NY.

MARAKAS G.M., 1999. Decision support systems in the twenty-first century. Prentice Hall, Upper Saddle River, NJ, USA.

MATTHIES M., GIUPPONI C., OSTENDORF B., 2007. Environmental decision support systems: current issues, methods and tools. Environ Modell Softw 22, 123-127.

McBRATNEY A.B., DE GRUIJTER J.J., 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. Eur J Soil Sci 43, 159-175.

McBRATNEY A.B., ODEH I.O., 1997. Application of fuzzy-sets in soil science, fuzzy-logic, fuzzy measurements and fuzzy decisions. Geoderma 77, 85-113.

MEDINA J.M., PONS O., VILA M.A., 1994. GEFRED: a generalized model for fuzzy relational databases. Inform Sciences 77, 87-109.

PANIGRAHI S., 1998. Neuro-fuzzy systems: applications and potential in biology and agriculture. Artif Intell Applic 12, 83-95.

PÉREZ-PUJALTE A., 1980. Mapa de suelos de la provincia de Granada, Escala 1:200000. Estación Experimental del Zaidín, CSIC, Granada, Spain. [In Spanish].

PETRIDIS V., KABURLASOS V.G., 2003. FINkNN: a fuzzy interval number k-nearest neighbour classifier for prediction of sugar production from populations of samples. J Mach Learn Res 4, 17-37.

POCH M., COMAS J., RODRÍGUEZ-RODA I., SÁNCHEZ-MARRÉ M., CORTÉS U., 2004. Designing and building real environmental decision support systems. Environ Modell Softw 19, 857-873.

RODRÍGUEZ A., BERBEL J., RUIZ P., 1998. Metodología para el análisis de la toma de decisiones de los agricultores. Ministerio de Agricultura, Pesca y Alimentación, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain. [In Spanish].

SMITH C.S., McDONALD G.T., THWAITES R.N., 2000. TIM: assessing the sustainability of agricultural land management. J Environ Manage 60, 267-288.

SYS I.C., VAN RANST E., DEBAVEYE I.J., 1991. Land evaluation. Agricultural Publications 7, Ghent University, Belgium.

THRUPP L.A., 1989. Legitimizing local knowledge: from displacement to empowerment for third world people. Agr Hum Values 6, 13-24.

WARREN D.M., 1989. Linking scientific and indigenous agricultural systems. In: The transformation of international agricultural research and development (Compton J.L., ed). Lynne Rienner Publishers, Boulder. pp. 153-170.

WEBSTER R., 1977. Quantitative and numerical methods in soil classification and survey. Clarendon Press, Oxford.

WEBSTER R., OLIVER M.A., 1990. Statistical methods in soil and land resource survey. Oxford Univ Press, NY.

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