EX.S.I. - AN EXPERT SYSTEM FOR THE DETECTION OF ACTUALLY IRRIGATED SURFACES

Giovanni Narciso, Pierpaolo Manzini, Stefano Lorenzini, Livio Rossi
ITA Consortium

ABSTRACT

On behalf the Italian Ministry of Agriculture (MiRAAF), the ITA Consortium has implemented the prototype of an Expert System for the estimation and mapping of actually irrigated surfaces. The study was conducted over 3 Regions of Italy, representing different climatic and agricultural environments: Puglia, Umbria and Veneto. The system represents the first, and up to date (1993), the only application to irrigation topics, in Italy of remote sensing techniques and expert analysis integrated within a Geographic Information System. The application required the construction of dedicated cartographic and alphanumeric data bases reporting the characteristics of the territory. The System provides as output a cartographic coverage of the study area reporting the actually irrigated surfaces and it allows the production of statistics on irrigation through the combination of thematic information with administrative boundaries and land use maps.

1 Introduction

The supply of water for all human and productive uses is an increasingly critical issue. The resource is lacking world-wide due to various causes among which, probably most important, is the increasing antropic pressure on the environment. This is particularly true for irrigation in agriculture, where environmental issues are highly conflictual with the objective of achieving the highest possible yields, also in consideration of the high costs of supply, transport and distribution of water.

Water is frequently either misused or unavailable; when the resource is available, distributed amounts may exceed what is necessary to achieve target yields or, the resource may be lacking when and where it is needed. These considerations are important when performing medium to long term planning of investments in irrigation infrastructures and it is not a rare event, that the main cultivated crops do not really require the sort of water inputs which are foreseen by the planner or that particular crops requiring relevant amounts of water, are cultivated where reservoirs are a limiting factor to production.

The availability of "reliable territorial information" and the robust knowledge of what is the "true" status of resources, should be the base on which decision makers direct their action, allocating resources where they are really needed. In the field of irrigation especially, the establishment of sound policies should rely on a complete knowledge of the yearly use of water in terms of requirements, amounts and interested surfaces, in one word "statistics".

Coming to the detail of the Italian situation, adequate information on the use of irrigation water is not yet available to the Authorities. The only existing data, comes from the 10 year National Agricultural Census, reporting the potentially irrigated surfaces at municipality level, without any further detail. Conscious of this, the Italian Ministry of Agriculture (MiRAAF), through the ITA Consortium (Consorzio Italiano per il Telerilevamento in Agricultura) decided to promote the study of a methodology for the timely and reliable detection of actually irrigated surfaces, possibly using satellite remote sensing as a primary source of data and capable of producing yearly statistics and maps. The Administration encouraged the development of all the possible sinergies with existing data bases and collection procedures (e.g. AGRIT Program for agricultural inventories),the integration of procedures in a GIS environment and the use of advanced means of data analysis. These indications led the ITA Consortium to choose an Expert System (EX.S.I.) approach to the topic.

The study was conducted over 3 administrative Regions representing different climatic and agricultural environments: Puglia, Umbria and Veneto and the analysis was targeted on the 1991 irrigation season (June - September) concerning mainly summer crops.

1.1 Motivations for the use of an expert approach

Main objective of the project is to assess the feasibility and effectiveness of a system addressed to estimate actually irrigated areas, based on Geographic Information Systems (G.I.S.) and remote sensing techniques.

This topic calls for the use of integrated survey tools, together with knowledge on the area and on the conditions that govern agricultural practices. Said information is not easily summarised or parameterized in a representative model. In fact it refers, at least for the "agronomic knowledge" component, to qualitative appraisal such as: opportunity of irrigation (for a given crop or phenological stage) and even the "traditions" of a given area.

Owing to the peculiarity of a system which aims at identifying irrigation as a land use class and the distinctive features of the examined field of application, artificial intelligence techniques and tools have been selected for development of a real Expert System.

The effectiveness of Expert Systems in dealing with non-conventional topics has been extensively demonstrated. and should be adopted in this particular case, for the following reasons:

- presence of multiple, differentiated and distributed sources of knowledge;
- presence of experience to be examined, analysed, organised and capitalised;
- presence of incomplete and/or inconsistent data;
- presence of varying topics that call for definition of deductive models, whose utilisation must include the presence of a strong heuristic component;
- need to adopt incremental prototyping, by examining and confirming knowledge, step by step, under dynamic evolution of scenario data.

Among the benefits of adopting an Expert Systems approach for project implementation are:

- they allow the expert to acquire greater knowledge on the relations between his area of skill and other types of information;
- the expert can participate to the analysis directly;

The complexity of agriculture and irrigation in Italy calls for a type of analysis that rules out deterministic approaches. Thus, heuristic solutions based on fundamental knowledge of the problem can be adopted, and then be appropriately enhanced and adjusted in time. The ITA Consortium considered this approach as the most suitable for the solution of this problem.

2 Data bases

The knowledge base for the System relies primarily on satellite data and the experience acquired so far demonstrates the need to support remote sensing information with "other data" only part of which can be defined as ancillary. More correctly all the data sets have to be considered with equal dignity in the process of inferring a surface feature such as irrigated crop canopy. In defining EX.SI. quite divers data bases were collected, relying for the necessary integration, on the structuring of the System itself

2.1 Ground survey:

The first action in the construction of the data base for EX.S.I. was the collection of intensive ground truth on 400 sites surveyed every 3 days throughout the 1991 summer. The sites are a sub-set of the sample normally used in the yearly enumeration for national agricultural statistics (AGRIT Program) and are around 50 Ha in size. They were mapped for ground cover and a number of features were monitored, most important of which was irrigation.

The collected data was digitised and it was structured in 2 data bases reporting the features monitored on the sites for each 3 day survey (land use, phenology, physiology, soil, irrigation, etc.) and the geometry of the fields in vectorial format.

2.2 Remote sensing data

Earth observation satellites are capable of supplying direct, repeated and synoptic information on the surface but using remote observation it is not an easy task to detect the exact moment of application of water as the timing of irrigation is subject to a number of random conditioning factors. The practice though has relevant effects on the vegetation canopy and these are quite evident both in absolute and relative terms. Irrigated plots generally appear in good physiological conditions and clearly stand out among non irrigated ones. A multitemporal approach was then chosen even though it is strong the need for reliable ground reference data and the identification of proper spectral variables connected to the feature.

The Thematic Mapper sensor mounted on the LANDSAT-5 satellite provided the imagery used for the study and as the areas of interest are covered by various scenes, the sequence of images in Table 1 was acquired.

Table 1

Region	LANDSAT TM - Track_Frame	Acquisition dates (1991)
Puglia	187-32	19/07, 04/08, 05/09, 21/09
	188-31	10/07, 26/07/11/08, 12/09
	188-32	10/07, 26/07, 11/08, 11/08
Umbria	191-30	16/08, 01/09, 07/09
Veneto	192-28	06/07, 07/08
	192-29	06/07, 07/0/

Irrigation is usually performed at times when Evapotraspiration is at its minimum, that is either before of after sunset and this modifies the thermal conditions of the canopy. For this reason a set of night passages, and in particular TM Band 6 data (10.4 - 12.5 μ), was also acquired. This data has a resolution of 120*120 m and is not exactly geometrically coincident with day time visible bands, all the same it has been considered useful for the "expert analysis" of a surface features as irrigation.

The sequence of acquired images is shown in Table 2.

Table 2.

Region	LANDSAT TM - Track_Frame	Acquisition dates (1991)
Puglia	50-212	05/09, 05/09
	50-213	19/07, 04/07, 12/09
	51-212	11/08, 12/09
Umbria	not available	
Veneto	50-216	01/07
	50-215	10/09
	53-216	10/09

Preprocessing:

As the images were acquired in different periods, in order to analyse the spectral evolution of the observed surfaces as related to irrigation, data has to be comparable. This requires both radiometric and atmospheric pre-processing:

- *Radiometric and Atmospheric Correction*: For the purpose, the normal *gain* and *offset* corrections suggest by the data provider were performed. Following a specific procedure was applied, integrated by data deriving from the *LOWTRAN* model.
- *Geometric Correction* : The Geometric Correction of the images was routinely performed with the use of Ground Control Points (GCP) and applying a nearest neighbour resampling method. This procedure was applied to both day and night imagery

2.3 Land Cover classification

Land cover classification is at the base of the whole methodology. Recognising the crop type, together with the spectral features, ancillary topographic and meteorological data and the agronomic knowledge base, allows the construction of a dynamic deduction process which leads to determine an “irrigation “probability”.

This “probability” can be considered “certainty” for particular crops or land covers in particular environmental configurations. For example *“tomato is surely irrigated in August in the Puglia region”* as it would not be there otherwise, or *“a natural forest is surely not irrigated”. The only source of uncertainty in these cases* is given by the accuracy of the classification, which can be measured, through a ”confusion matrix”.

Satellite images were processed to derive 2 band composites of “GREENNESS” and “BRIGHTNESS” indices. The spectral signatures for the various crops were then selected with an automatic “supervised” procedure using the ground truth samples as training areas. This methodology was adopted for the multitemporal coverage obtaining for each class a “cluster” of signatures which was used to run a “minimum distance” classification algorithm.

Once the images were classified, for each frame a confusion matrix was produced on the base of the ground survey. The global accuracy of the classification, computed as a ratio between the number of “correct” pixels and the total number of pixels, resulted within 60% *and 70% for the various frames. Considering that some 25 - 30 crops were classified*, this figure can be considered acceptable.

Since on the study area there was a certain overlap of the frames, once the classification was completed, a mosaic of the images was recomposed, choosing for the overlapping portions, the frame with the “best” confusion matrix.

The classified images have all been filtered in order to eliminate isolated pixels, which are essentially noise at the project scale of analysis. For this operation a rectangular (3*3) pixels filter was applied and with two sequential passages, a set of homogeneous units with a minimum size of 0.36 Ha (4 pixels) was created. This units (*Blob*) can be roughly considered as fields, even though there is no real connection to the actual geometry on the ground and are the base spatial element of the expert analysis.

The classification methodology is shown in the following flow-chart Fig. 1:

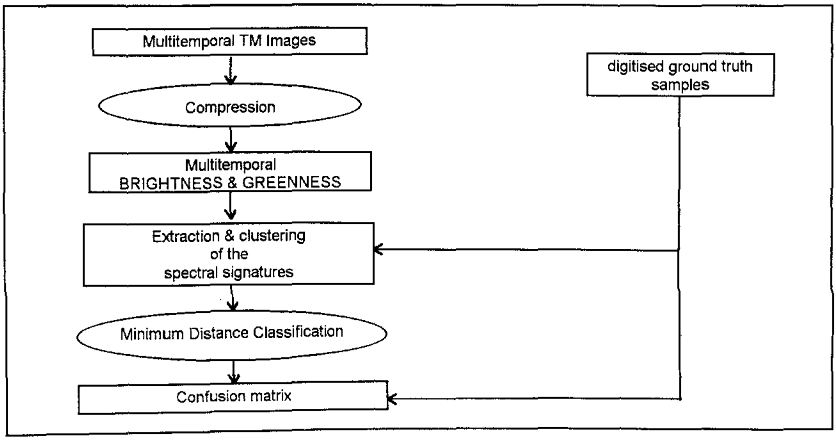

Fig 1. Flow chart of the classification procedure

2.4 Vegetation Indices

If for certain crop/environment patterns, the application of irrigation can be considered "sure", this is not the case for most of the crops. Irrigation, though, has quite perceptible effects on vegetation and one can reasonably say that, good crop conditions (or their improvement) in an unfavourable environment, can be reasonably charged to irrigation. Remote sensing is an efficient mean to determine crop conditions through what are commonly defined "Vegetation Indices". These are band combinations based on the reflective properties of the canopy and their validity is widely demonstrated in literature.

In the EX.S.I. project, a number of vegetation indices were tested in order to determine the ones which best outline irrigation. The reference index was the NDVI and apart from most widely used normalised difference of Bands 3 and 4, some additional combinations with Bands 5 and 7 of TM were tested. The Indices have the following formulation:

$$NDVI_{(1,2,3)} = \left(\frac{TM(4,5,7) - TM3}{TM(4,5,7) + TM3} + 0.5 \right) * 100$$

Where: *NDVI* = Normalised Differences Vegetation Index

TM3 = Spectral radiance for Band 3 of TM

TM(4,5,7) = Spectral radiance for Band 4,5 and 7 of TM

A further Index was tested, the VMI (Vegetation Moisture Index) as the difference of the spectral radiance of Bands 5 and 7.

$$MI = (TM5 - TM7) * 100$$

Where: *VMI* = Vegetation Moisture Index

TM5 = Spectral radiance for Band 5 of TM

TM7 = Spectral radiance for Band 7of TM

The thermal band of TM (Band 6) was also used in the analysis, using both daytime and night time passes. Two indices were derived:

- Thermal Radiance (TR) when only the night passage was available. Data was transformed to temperature at the sensor, under the hypothesis of unitary emissivity and validity of pre-launch calibration constant.
- Apparent Thermal Inertia (ATI) was derived when both day and night images were available on the same date:

2.5 Data analysis

The vegetation indices were analysed versus the ground truth to choose the most adequate for a specific crops and to define specific threshold values of irrigation. Data was available for every third day so it was possible to have an almost complete description of the sample sites in coincidence with the satellite overpass. Using the geometric layout of the sites, the indices were extracted from the images for each surveyed field and stored as maximum, minimum, average, mode and median, *in a specific archive together with the* quantitative/qualitative description of the field.

The analysis followed a recursive procedure in order to determine for each combination of crop and/or phenological phase the index which significantly discriminated the irrigation event. The procedure was applied for each date of acquisition and for each frame.

The vegetation indices were compared to the Boolean variable "*IRRIGATED YES/NO*" *and a* Pearson correlation matrix was used to operate a selection of the best performing index. This operation was followed by the selection of an irrigation threshold on the frequency distribution of the selected index per homogeneous <crop/phenological phase> configurations using a Chi-square method.

The methodology is summarised in the following scheme (Fig 2) :

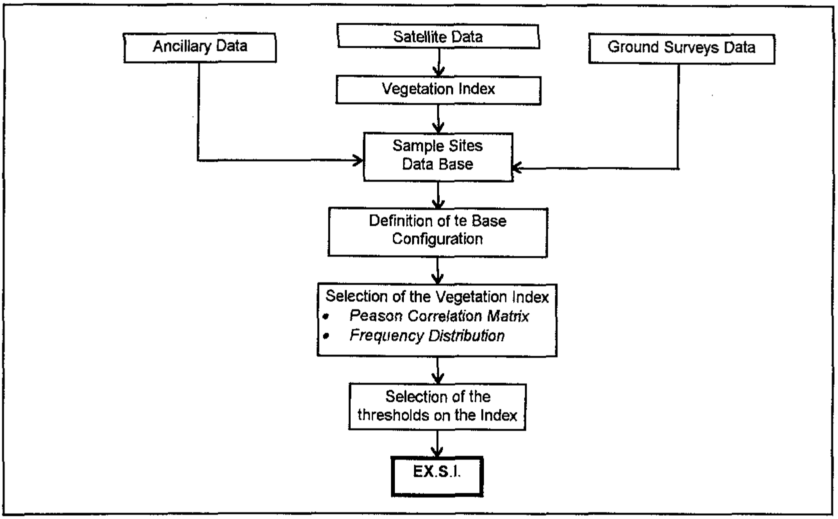

Fig 2. Flow chart of the vegetation indices analysis

2.6 Thematic cartography;

Essential component of the data base is a set of cartographic information of the 3 Regions, which were managed and analysed during the various phases of the project using GIS technology. Following is a brief description of the data set.

General land use: Land use information has been derived from a 1:25.000 scale Land Cover Map prepared by the ITA Consortium on behalf of the National Board of Statistics (ISTAT) and the Ministry of Agriculture (MiRAAF).

This is the most updated document available in Italy and was derived from the interpretation of a national coverage of SPOT-XS scenes. The map includes 6 classes of land cover:

- Agriculture - Arable Land
- Agriculture - Permanent Crops
- Forest
- Urban areas
- Non vegetated areas
- Surface water

The first two classes were use to outline the areas of interest for the project.

Potentially Irrigated Areas: This document (1:100.000 scale) outlines the administrative boundaries of the irrigation schemes and was used to furtherly define the interest areas.

Digital Terrain Model (DTM): From a national coverage of digitised altitude points, two documents were derived:

- an altitude map;
- a slope map.

Available Soil Water Capacity Map: This document was derived from a regional sample of soil profiles. The cartographic information, reports the limits of homogeneous units and is linked to a data base with the specific soil characteristics of the single units.

2.7 Ancillary data

Meteorological data: Rain and other meteorological variables have a relevant impact on the irrigation practice and they have to be accounted for in the expert analysis. The information was derived from the data base of a Spectral Agrometeorological Model (SAM) for yield forecasting (MiRAAF) where data is referred to a national grid with cell size of 8*8 km.

Phenological data: The SAM also provided the specific information for the phenological evolution of the main crops. A biometeorological module (BIOM) defines the various phenological phases as a function of meteorological inputs on the stated grid base.

Statistics on land use: On the base of the ground survey, a series of statistics was derived in order to describe the connection between the cultivation practices of the crops and irrigation:

- percentage of cultivated surface per crop
- percentage of irrigated surface per crop
- percentage of irrigated surface per crop and per phenological phase

Other statistics were derived by matching the survey information with the available cartographic data bases (height, slope, available water capacity of the soil, potentially irrigated areas, etc.)

Agronomic features of the crops: The previous information can be considered "quantitative" and it has to be integrated with was is commonly outlined as "agronomic knowledge". This outlines all those generic and/or qualitative information which derive from experience and are capable of connecting and explaining relations between environmental items, crop characteristics and the irrigation practice.

Agronomic knowledge was formalised by structuring "experience" in a series of rules which, even, though not exhaustively, describe, in an homogeneous manner the connections between an "environmental configuration" and irrigation.

Rule definition has been organised for the specific crops by defining an "agronomic form" where the quantitative knowledge items are given a ranking and a value, expressing the relevance of the item in determining irrigation. For some particular items (e.g. rainfall, height, slope, etc.) it was possible to define quantitative ranges and limits, for others the connection could not be rigorously established and the association was expressed by a mere qualitative judgement of significance, assigning a relative weight to the items itself.

Concepts related to the agronomic knowledge were not limited to the described form but played a primary role ad various levels of development of the System.

3 Characteristics of the Expert System

Expert Systems are organised in various modules, whose role is accurately defined during implementation of the system. The following components are always present and can be singled out when dealing with knowledge-based systems.

Knowledge base (KB):The KB is a module that acquires knowledge on the field of application, expressed as basic considerations (facts) and their interrelationship (rules). One of the highlights of KB is that it separates knowledge on the field of application from other types of knowledge (e.g. knowledge on communication with users).

Inferential Engine: The Inferential Engine is a module that draws on the KB for reasoning. It selects various types of rules among the options for achieving solutions.

User Interface: The User Interface provides machine-user interaction in such a way that the user can establish a real dialogue with the system. If this module is used in an expert mode, the interactions between User and System gradually become more guided, and thus parameters and knowledge are easier to handle.

As shown in the Figure 1, EX.S.I. contains all the above-mentioned concepts, adjusted specifically for the examined topic.

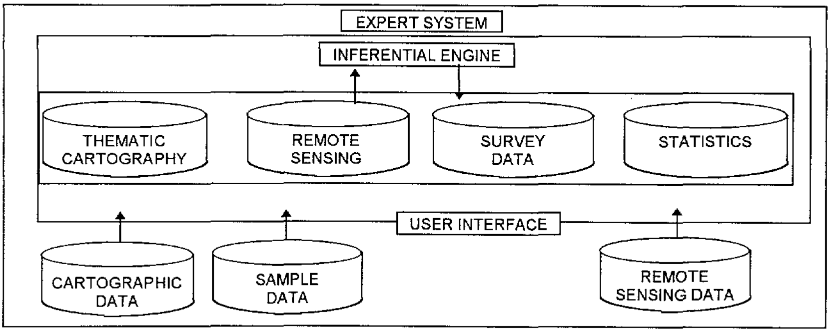

Fig 3. Scheme of EX.S.I.

The Inferential Engine and the User Interface of EX.S.I. are rather anomalous compared to other systems. In fact, the Inferential Motor is specifically addressed to classification of irrigation, whereas the User Interface does not enable dialogue with the user but permits effective interaction for calibration of the system's parameters.

The KB plays a decisive role in the system; it contains all the rules derived from the analysis of ancillary data that describe the examined area (maps, remotely-sensed data, ground survey). In addition, it incorporates knowledge elicited from expert correlation's of non-homogeneous data and their statistical processing. Thus, acquisition of knowledge is the fundamental requisite for development of the components forming EX.S.I., and to define an overall approach for system construction.

Knowledge acquisition proceeds similarly to software development although development of an Expert System is essentially evolutionary and leads to progressive fulfilment of the system. In particular, the knowledge construction stage (and consequently, construction of KB) develops as follows (Fig. 4) :

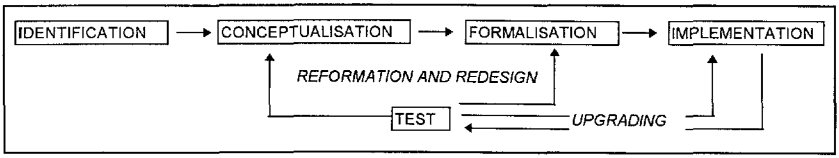

Fig 4. Scheme of knowledge acquisition.

Preliminary survey on field of application :This preliminary stage, based on interviews and collection of detailed information, is addressed to acquisition of knowledge on irrigation in 3 study areas, as well as identification of all data and parameters that could be useful in finding a solution to the problem.

Detailed investigation of the topic with identification of initial knowledge: Acquisition of initial knowledge requires the preparation of "agronomic forms" and analysis of remotely sensed data (classification, vegetation indices, etc.).

Progressive upgrading of the basic knowledge: By means of various investigations and by varying input weights and parameters, the KB is progressively upgraded, and a more reliable version of the system is achieved. The activities that led to fulfilment of the EX.S.I. are:

- analysis of data and their interrelations;
- statistical analysis.

4 General description of the System

The ancillary data and the information stored in the KB are structured in files which the system considers as "objects" and starting from the land cover classification, a probability of irrigation is determined. In Figure 5 the steps are briefly described:

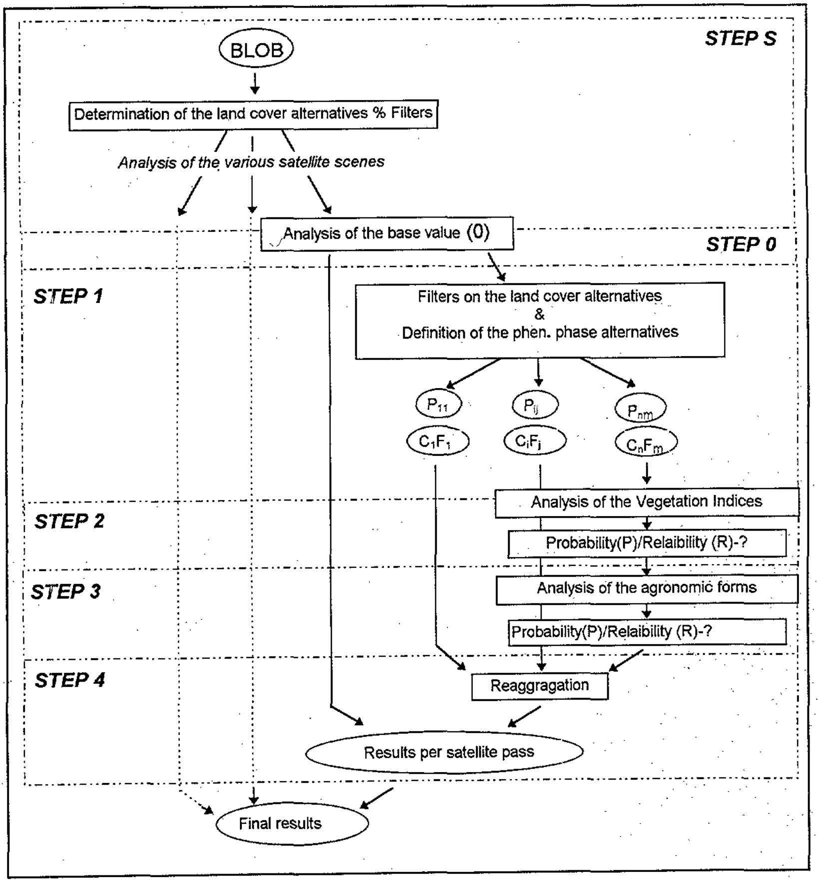

Fig 5. Scheme of EX.S.I..

Step S: This preliminary step examines land cover alternatives,(expressed as percentages in the confusion matrix). Alternatives which are judged not plausible are automatically rejected. The process is repeated for all the classified images.

Step 0: The configurations (crops, height, slope, etc.) which do not require an analysis to infer irrigation probability are picked and assigned either to *Irrigated* or *Non-Irrigated* classes.

Step 1: A number of <crop/phenological phase> sets are derived from the alternatives chosen at *Step S*, assigning a probability of irrigation to each specific Blob of the classifications. The analysis precedes up to *Step 4*, for each alternative and for each image. At this stage, as the single Blob may change its configuration on the multitemporal passages, a filter is applied to exclude non plausible configurations. Each set <crop/phenological phase> is classified as a function of the following considerations (coded in the KB):

- Surely not irrigated (*Step 2* and *3* are skipped);
- Surely irrigated (*Step 2* and *3* are skipped);
- No sure assignment can be made

Step 2: This step determines a probability of irrigation for the analysed Blob and evaluates the reliability of the hypothesis. The probability is established analysing the vegetation index of the Blob and the related thresholds of irrigation. At this stage the morphological (height and slope) configuration of the Blob is analysed together with meteorological variables (rainfall) at the moment of the satellite overpass in order to define possible conflictual situations. If no index or threshold are available the analysis proceeds to *Step 3*.

Step 3: At this step the KB and the crop agronomic forms are analysed to refine the reliability of the hypothesis produced at the previous step.

Step 4: The results on the various alternatives of <crop/phenological phase> are reaggregated combining probability and reliability. The criteria is quite simple and the best outputs for the single Blob, at the various dates, are selected.

5 Results

Irrigation map: As final result of the expert analysis, the System provides a value of probability of irrigation for each single territorial unit (Blob). It is then possible to recode the mosaic of classified images, covering the study regions and create an "Irrigation Probability Map", with classes ranging from 1 to 100.

This document, though rich in informative contents is scarcely comprehensible to a non acquainted user. A slicing of the value was then performed using, as reference, information on irrigation derived from the National Agricultural Census (information not used for the Expert System). Comparing the Census surfaces with surfaces on the map at various level of irrigation probability, a specific threshold was determined for each of the 3 Regions and a binary map was produced with two clear cut classes: "Irrigated" and "Not Irrigated".

Statistics: Matching the produced map with the administrative boundaries of the municipalities and the national land cover map it was possible to derive a series of statistics of the actually irrigated surfaces and provide the Administration with a comprehensive bulletin. An example of the bulletin for 4 municipalities in the Puglia region is give in Table 3.

ISTAT CODE	Municipality	Tot. Sur. (Ha)	SAU (Ha)	Agrarian region (code)	Alt. class (code)	Pot.. Irr. Sur. (Ha)	Pot . Irr. Sur./SAU (Ha)	Act.. Irr. Sur. (Ha)	Act.. Irr. Sur./SAU (Ha)
71001	**ACCADIA**	3071	2553.3	1	1	-	-	-	-
71002	**ALBERONA**	4969	4411.9	3	3	-	-	8.91	0.20
71003	**ANZANO DI P.**	1105	1039.2	1	1	-	-	-	-
71004	**APRICENA**	17243	12522	7	5	7340.9	58.62	1358.82	10.85

6 Conclusion

The illustrated study was essentially a feasibility exercise the outcome of it though, proved the maturity for operationality of the adopted technologies as it succeeded in completing the whole cycle from data collection to data generation.

As for all prototypes a great amount of effort has still to be placed in engineering in order to create a properly structured Expert System. At the present moment (1995) a more advanced program is bound to start, interesting a total of 8 regions in Southern Italy.