

Computer-based methods in environmental data organization and management

Abstract

The protection of the environment is one of the main problems the mankind faces today. Recent advances in computer technology have assisted environmental studies towards this direction, in various scientific approaches. In this study, an attempt was undertaken on the integration of various environmental data (mainly agricultural) into databases with database management facilities and data analysis methods, using an interactive information system created for this purpose and called *AgroModel*. The aim was to provide a proper educational and friendly system for a more efficient integration and manipulation of huge environmental data in order to be used for educational and research purposes.

Arising out of a need for a better organization and exploitation of environmental knowledge, an integrated interactive computer system for environmental (plant & animal) education and research purposes was developed and used efficiently in local area research cases. The system provides build-in tools as well as interfaces to software packages for mathematical and statistical analysis. Being in such an integrated environment, the researcher could undertake any type of statistical analysis from the simplest descriptive statistics to the most sophisticated statistical approaches for the derivation of important parameters for scientific and practical applications.

Keywords: Integrated computer system, huge environmental data

Volume 8 Issue 3 - 2019

Stergios Tzortzios, GK Adam, N Dalezios

University of Thessaly, Greece

Correspondence: Stergios Tzortzios, University of Thessaly, Greece, Email stzortzios56@gmail.com

Received: April 23, 2019 | **Published:** May 02, 2019

Introduction

A number of computer products have been developed and applied in environmental education and research in general.¹ However, there are still a lot of environmental issues that need to be tackled further in order to provide solutions, particularly in the organization and assessment of the various scientific findings. In this direction, an interactive application environment, called *AgroModel*,² was constructed and used as an integrated computing platform for efficient organization and management mainly of agricultural data. The basic idea was to use an integrated environment, where various data analysis tools could be applied in order to extract and propose solutions to certain case environmental problems.

The system is built using object-oriented software (integration of Visual Basic, SQL and web development languages HTML)^{3,4} and incorporating certain statistical analysis packages used also for advanced programming and work with the system. *Agro Model* sessions administrate a huge database, called AgroDB, created on a relational model scheme, where various plant and animal descriptive and research data were stored. The database is under continuous improvement and updates since new data and findings could be continuously added.

The application environment

The overall application work was carried out using the above mentioned integrated environment *Agro Model*. The development of this application environment was based mainly upon the use of object-oriented and visual development tools and techniques,⁵ and the use of open architecture technology drivers and methods -ODBC interface, SQL- to interact initially with Microsoft Access and Excel databases and certain statistical packages later on, such as SPSS, on a

Windows operating system platform. In particular, software modules were created as VBasic modules scripts and SQL queries in order to facilitate the communication of system components and execution of internal functions and procedures.

The idea was to develop and use an interactive and friendly environment for agricultural data management, where data analysis (e.g. statistical) and experimentation, as well as further research on contemporary data analysis techniques, could easily be contacted. Some of the basic system tools and facilities the environment provides include database management and maintenance, data analysis and advanced data management involving some type of programming.

The environment of *Agro Model* currently provides the following basic tools for agricultural data management:

- Database basic data manipulation (retrieval, update, filter, report, queries, etc.)
- Data analysis (e.g. statistical, etc.) through build-in functions, spreadsheet functions, etc.
- Advanced data management (based upon logical reasoning, SQL queries, etc.)

The data analysis *tool* provides access to spreadsheet functions (e.g. Excel), where the user can either perform data analysis directly on a spreadsheet worksheet, or work with data analysis tools and functions within the application environment. Some of the basic data analysis facilities include: mathematical analysis and statistical analysis (i.e. mean and standard deviation, standard distributions).

The advanced data management tool currently provides a small number of logical functions, based mainly on quantitative techniques, for the generation of any required new information derived from the

existing database; nevertheless, this functionality is under continuous development. A general schematic view of this environment is presented in Figure 1.

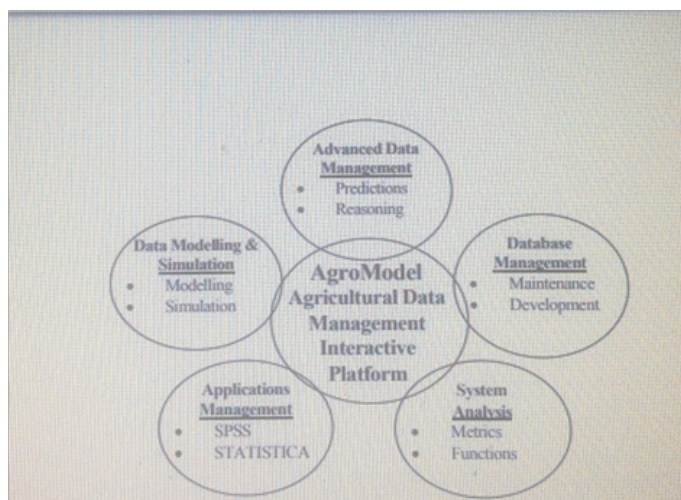


Figure 1 The AgroModel basic computer environment.

The application's user interface design is user-centered (user-friendly), based upon system objects (tables, forms, frames, etc.) through which the user could easily communicate and interact with the functions and procedures provided within the environment to obtain the required information from the database.

Database organization

The overall agricultural data used in our case studies are organized as a database and manipulated with AgroModel database management facilities. The database has been currently filled with agricultural information about plant and animal species (i.e. sugar beet and cattle breeds data), collected over a long period of time and has started taking into consideration a number of relative environmental effects and factors (various meteorological data) stored in any level of measurements within relational tables.

However, the system comprises a flexible structure that is under continuous improvement and evolution. At this time it is under further integration with a geographical data analysis tool (*GisTool*) used for experimental research in GIS.⁶ As a result, further environmental information is being added based on local topological characteristics (e.g. populations, areas, land usage, etc.), grouped in relational tables and forming a large amount of parameters some of which constitute the basic ones in further logical and arithmetic processing. Since geographic features are usually identified by multi-dimensional tables the logical geographic data structure is capable to provide maximum efficiency in queries and retrievals executed during data processing.

The final platform produced seems that gives another insight on the environmental issues, assisting further students and researchers in their work of understanding and experimenting with environmental information.

Database structure

The Agricultural Database (AgroDB) as a whole was designed in the form of interrelated tables, grouped together according to the specific type of the agricultural data (i.e. scientific field, plant species, breed, etc.), which describes how the data (entities and their

attributes) within the database are related at all levels (linked or related to other data inside the database). Taking into consideration that relational databases are formed on the basis of a mathematical reasoning (relational calculus, relational queries), while hierarchical and network databases have almost no mathematical footing, a relational database schema was selected and used in the construction.

A database scheme structure representation is given in Figure 2. As it can be seen, the overall organization of the agricultural database is based on a relational database model, where each specific entity of agricultural data is structured logically as a table data structure and related to other data subsets through the appropriate key data fields. For representing these table relations, a shorthand notation is used, where each data attribute (table field) is contained in a list separated by commas, with the primary key underlined and the name of the table to the left of the brackets.

[Region (RegionId, StationId,...)

Station (StationId, PrefectId, Year, Season,...)

Prefectures (measurements of temperature, rain level), etc]

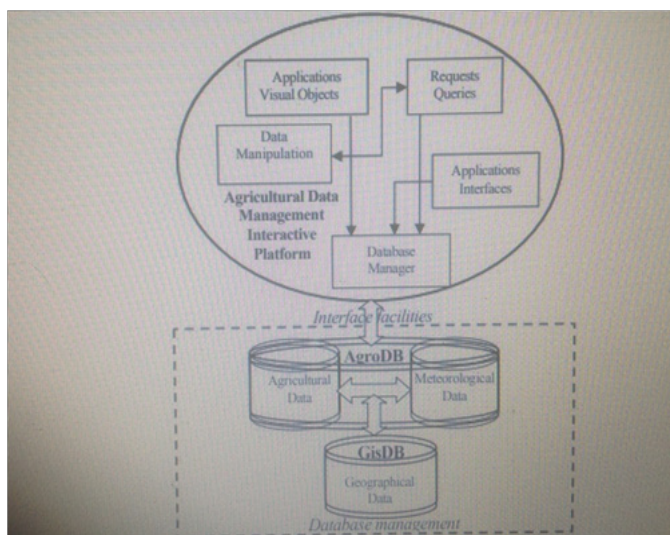


Figure 2 A scheme structure of the database.

As we can see from the above, each table has a unique property field, as a primary key, e.g. *RegionId* acts as a unique identifier for a particular record of information within *Regions* table. In other words, all these attributes within the entity *Regions* table are functionally dependent (have a unique association) with the primary key, *RegionId*. This approach simplifies the management of the relational tables, by reducing their relations to the simplest forms, and therefore making them easier to handle, and ensure that data could be processed more efficiently (e.g. queries could be carried out more quickly).

Data manipulation

The management of the above database structure could be easily performed using AgroModel's basic database management facilities, or in a more advanced and constructive way (more sophisticated retrievals), by using specific SQL queries provided for this purpose. A partial view of such data manipulation is shown in Figure 3.

Finally, this database structure also allows applications (e.g. Excel spreadsheets) and high-level languages (e.g. Visual Basic scripts, SQL queries) to link to the data in such a way that the whole structure of the database is transparent to such or any other applications according

to the users requests. At the top level, the application system of AgroModel controls and coordinates the entire user interaction and requirements (educational or research), towards the agricultural database, through a number of facilities and tools provided for this purpose.

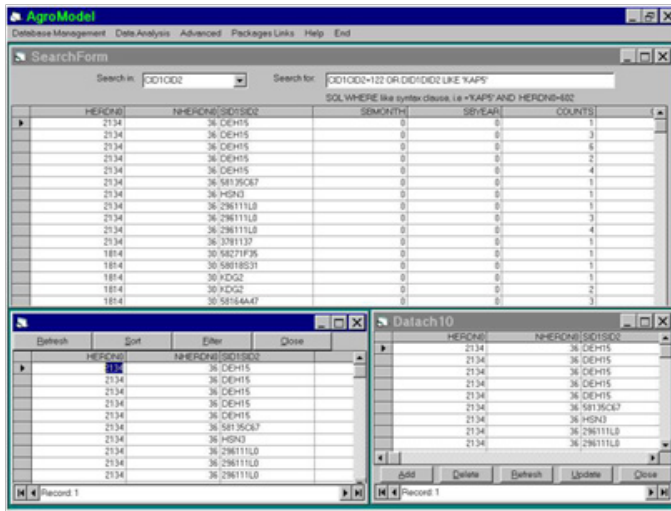


Figure 3 SQL-based database management window.

Results and discussion

It was important to find the tools to describe and analyze such agricultural data structures (i.e. cattle breeds and their characteristics), in order to specify and select the most adequate scheme, without an in-depth requirement for programming skills. It was also necessary to be flexible enough to allow easily modifications of the given data structure according to any new requirements.

Case study: most optimum selection of specific cattle breed

In our case study, we used only a subset of agricultural data (of Charolais Breed) extracted from the database to work with, as an example of farmer’s level specific interest. Exploiting the utilities afford mentioned, the following is a simple example of a specific interest at the level of farmer. It refers to a selected Charolais specific data file from the agricultural database and more particularly to the most optimum selection of dams according to specific parameters.

In this particular example, a certain range of calves characteristics was traced, starting from their birth up to 200-days and 400-days weight, taking into consideration various calf specific parameters and environmental factors and the information referring to their herds, parents (sires and dams), etc. During the working session, a model was produced of about 100 out of 1330 calves’ cases that were analyzed using mixed procedures based on classical statistical methods and internal tool procedures. The modeled session included the specific properties of herds, sires, dams and calves (e.g. age, sex, weight, etc.), and their interactions. In particular, in that specific case it was estimated that about 50% of dams found to have superiority (class) definition parameters well above the overall average.

A considerable part of the above animal scheme was also compared and justified against the results obtained through specific measurements carried out using internal to the model SQL queries, such as the following: SELECT cid1cid2, sid1sid2, did1did2, nherdn0,

hszse, hcsla, sclsa, dclsa, year, csex, twin,

w3, aw3, hw3, w5, aw5

FROM charolais

WHERE hcsla>=3 and sclsa>=3 and dclsa>=3

AND cid1cid2 IN (select did1did2 from charolais);

A results report is presented in Figure 4, where the selected coded heifers (cid) are the most promising dams, according to interesting (genetic and/or environmental) parameters included in the database. The importance of this event is very well known that plays a crucial role in the farm’s welfare.⁷ Although the best performing animals do not necessarily have the best genetics, in this case we were able to predict and select which animals have superior performance based on genetic factors coming from their parents’ specific attributes, classes etc., and other environmental effects.

cid1cid2	sid1sid2	did1did2	nherdn0	hszse	hcsla	sclsa	dclsa	year	csex	twin	w3	aw3	hw3	w5	aw5	hw5
AHL3	AAE2	58238F4E	33	3	3	6	6	75	1	0	321	309	309	508	496	496
HSL3	AAE1	HSF14	15	3	5	5	3	75	1	0	231	272	272	494	504	504
AHN4	AAE1	58238F41	33	3	3	5	6	73	1	0	279	287	271	470	468	453
DEI2	AAE1	71019C22	19	6	3	5	3	73	1	0	324	327	327	508	517	517
HSM7	AAE1	71122F25	15	3	5	5	6	76	1	0	259	280	280	441	458	458
BRJ15	AAF1	58893F34	29	4	5	6	6	74	1	0	379	309	309	510	488	488
KFH22	BGE13	58028F23	26	1	4	5	5	72	1	0	268	287	263	367	410	386
KLN3	BPJ10	KLJ8	23	2	6	4	5	77	1	2	210	321	266	422	443	431
BRJ10	BRJ9	BGC6	29	4	5	5	6	73	1	0	340	298	298	524	492	492
CDL15	BRG6	GTH21	29	4	5	4	5	75	1	0	231	265	250	422	422	407
BRM12	BRH08	03163H12	29	4	5	4	4	76	1	0	191	245	236	434	419	410
BVM03	BVH03	03222F14	4	2	3	3	4	76	1	0	297	270	270	444	425	425
CW15	CWC16	CWE49	8	2	6	5	6	73	1	0	342	334	321	495	496	483
DEJ8	DEG2	58270F04	19	6	3	4	4	74	1	0	256	300	288	417	436	424
JFN9	DEH15	AMKJ3	7	3	4	4	5	77	1	0	248	283	272	464	471	460
AAW4	DEH15	FFI20	36	1	5	4	6	77	1	0	325	370	370	435	455	455
DEN3	DEI26	DEI7	19	6	3	4	6	77	1	0	27	302	285	510	524	507
LFN4	GHF6	LFY4	24	3	3	4	3	77	1	0	262	289	278	412	428	417
LFM1	GMP5	LFY1	24	3	3	3	6	76	1	0	261	300	283	386	406	389
HSM3	HSJ1	71108F06	15	3	5	5	6	76	1	0	227	294	294	454	502	502
HSM6	HSJ3	HSJ07	15	3	5	6	6	76	1	0	306	325	328	460	484	487

Figure 4 Best performing animals’ selection table report.

Overall performance results and summary

Being in a well-organized database the scientist-teacher or student– could attempt various data manipulations in order to derive interesting specific genetic or environmental parameters; to create new variables for various applications; to aggregate groups of certain purposes; and so on. The practical meaning of the users’ familiarization with such data handling utilities is obviously of great educational importance, because:

- It helps him in developing the necessary self-confidence in approaching the material in study,
- It offers the chance for a better understanding of the statistical material’s meaning as the unique source of any research study, and
- It contributes to the gradual development of the proper scientific mind as a pre-requisite of a possible more integrated form later on.

After all, going on to the stage of the statistical analysis level of learning, it could be supported that a system based on a more theoretical approach would be rather boring and not far from a negative result regarding applied biological sciences as it is that of agriculture. A system more practical, which suggests a straight involvement to the problem, will be much more encouraging and

because of this it will be expected to create satisfaction. Some years ago it was rather difficult – if not impossible – to undertake a such approach because of the complete lack of the proper equipment. It is however, incredible nowadays, with the enormous development of the science of computing statistics and its available utilities as important tools in the agricultural applications in the field of environmental science, to forget the exploitation suggested.

More often then, it is discovered that in the best case the real models are only partially in agreement with the original ideas or at least the ideas do not seem to be such as clear in their applied form always. Thus, a repeatable process starts being developed in which the original ideas change slightly according to the empiric findings, and a new analysis is undertaken for testing the new ideas, and so on, new ideas suggest new analysis. This repetitive process is going on with the hope of reaching the point of understanding the inter-relations reasons and factors as they are presented in the models of the ideas in the clearest way.

In this process of the inductive training the computer has become the most useful tool. First, effectively designed computer programs facilitate the manipulation of ideas and findings, making this process fast and tireless. Second, such programs running in high- speed computers produce an outbreak in statistical ability. Third, these programs have made it possible to test scientific theories based on huge number of variables, which was practically impossible to be handled some years ago. The whole job becomes particularly easier when the various programs, as applications, are constructed in the format of a unique system of conventions by which the user could interact with the programs. A well designed system could allow the user to execute a series of jobs in the least of the time spent for the manual's advice and data handling.

Conclusion

As a general conclusion of the results presented could be realized that the degree to which a user could exploit the *AgroModel* depends on his goal and his level of knowledge and experience. In cases of general database use, an elementarily trained farmer could accomplish his main managerial requirements; for a more educated researcher there are advanced tools, which could be exploited according to specific scientific purposes. We have addressed the problem of organizing and managing some of the important cattle breed characteristics, using a tool that utilizes computer-based methods and statistical tools, in order to deal efficiently with proper animal's selections of the most productive ones.

In particular, we were able to:

- a. Produce a reliable description of the specific cattle breed scheme based on SQL models.

- b. Provide flexibility in the manipulation of various cattle parameters in SQL query forms.
- c. Produce accurate results of cattle's behaviour comparing to physical representations.
- d. Reduce the cost of cattle's management by reducing the risk of taking wrong selections decisions.

This approach of the application system *AgroModel* among its general importance in the agricultural industry as a whole (education, research and production), seems to be a very useful tool in organizing the information regarding environmental data in general, while at the same time provides the utilities for the best exploitation of the knowledge gained up today in the field of agriculture. This application environment tends to be improved and incorporate further agricultural and environmental data in general, as a broader network database, in order to provide a fully integrated environment where various environmental data could be gathered for studies of scientific and practical purposes.

Acknowledgments

None.

Conflicts of interest

Author declares that there is no conflict of interest.

References

1. Riley J. *Statistical usage in biological sciences: main problems and solutions*. Mexico: Proc of the 5th Meeting of the International Biometric Society. Network for Central America and Caribbean; 1997.
2. Tzortzios S, Adam G. *Proper computer procedures (AgroModel) in plant and animal selection for research and educational purposes*. Moscow: Proc of the International Workshop On Information Technology in Agricultural Education; 1999.
3. Darwen H, Date JC. *Foundations for object/relational databases*. USA: Addison-Wesley Publishing Co; 1998.
4. Craig JC, Webb J. *Microsoft Visual basic 6.0 Developer's Workshop*. USA: Microsoft Press; 1998.
5. Law AM, Kelton WD. *Simulation Modelling and Analysis, McGraw- Hill International Editions. Industrial Engineering Series*. USA: McGraw-Hill; 1991.
6. Adam GK. *Gis Tool: A tool for and processing manipulation of geographical data*. Athens: Proc of the 1st National Conference in Geographical Information Systems, Greek GIS Society Press; 1999.
7. Tzortzios S. *Special biometrical tools in selecting groups for multiple traits*. Vienna: Proc of the 48th EAAP Meeting; 1997.