OENOSES 3 (1): 41-53, 1988

DAPROPHECO: A SPECIALIZED DATABASE SYSTEM FOR INTEGRATED STUDIES IN PLANT-HEMIPTERA COMMUNITIES

Panos Petrakis, Makrygianni 70, N. Ionia 142 35, Athens, Greece

Keywords: Database, Program, Models, Relational, System, Plants, Communities, Hemiptera

Abstract: A realization of a specialized database system is illustrated through examples from data processing in integrated studies on plant/Hemiptera communities. A formalized algebra on a set-of-files level is presented and its functionality on real data bodies is considered. Also, some peripheral works interfacing this system with related statistical packages are illustrated with previously defined file structures. It was found that advanced computerized management of large and multifiled data bodies are accessible even from an inexperienced user and this is done through a completely menu-driven usage of the package.

Introduction

Although ecological communities possess an enormously multifairious structure and expressed through an infinity of phenomena not readily perceptible, ecologists have developed efficient methods to retrieve data and store them up for subsequent analyses. Usually the methods construct their own "model" of this data set by arranging it in a single matrix. A subdivision of this single matrix into smaller submatrices is usually necessary if we are aiming to reduce the zero entries and simplify the original structure into more compact portions. But this operation has hidden disadvantages emerging when we try elementary works, such as searching for a species on a certain topographic position, deleting records of a certain time period, declaring new variables and storing data to their domain. Especially on integrated monitoring projects these operations are extremely time consuming even in the case of computerized manipulations by the most efficient retrieval algorithms operating on a single file level (Estabrook and Brill, 1969). But in this type of projects we face a much more difficult problem, that of the maintenance of a complex set of directly or indirectly interrelated files the managenent of which will be done by an appropriate system efficient in integrating all files in a data base. Additionally, this set of files has to be flexible and serve several different purposes when viwed by more than one persons who have distinctly allocated interests.

In the sequel, I discuss in brief the philosophy of an ideal database and my approach to this model of organized - more accurately reorganized - biological information, though the recently developed theory of relational view models (Codd 1970, 1982) and the operational realization of this theory (Astrahan *et al.*, 1976; Castro *et al.*, 1985).

Philosophy and semantics in the database model

My approach starts from the view point that all man

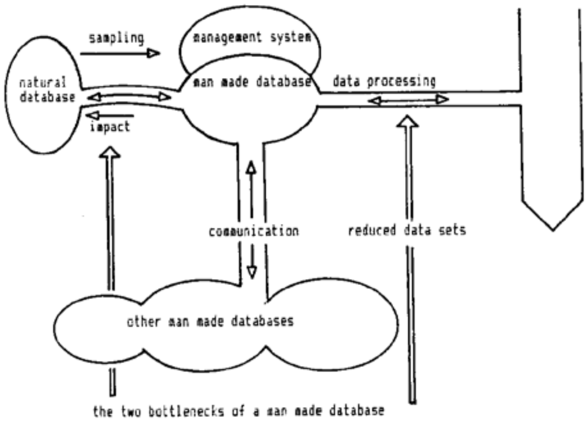


Fig. 1. Diagram for the illustration of information flow in ecological studies (explanations in the text).

made databases are subsets of natural databases which are variously named by community ecologists, namely communities or ecosystems. I accept the opinion that man made databases are the result of a definite small set of algebraic-like operations on one or more natural databases. These operations are called *projection* and *selection* respectively and will be formally defined in the sequel. In general, this results in an oversimplified model, with a certain loss of dimensionality and information. In spite of this oversimplification, the dimensionality of the reduced reality is still far beyond human perception. A natural reaction of human intelligence is the establishment of classes in which individual data sets could be alloted by a subjective (Braun-Blanquet, 1964) or an objective (among others Pielou, 1977; Orlóci, 1978; Greig-Smith, 1983; Goodall, 1986) procedure. This reaction is best summarized by the familiar concept of *classification*. Another type of this reaction includes the construction of a system of up to three axes creating thus a topological space in which data sets are placed as points in such a way that the maximum information on group structures is retained by the most

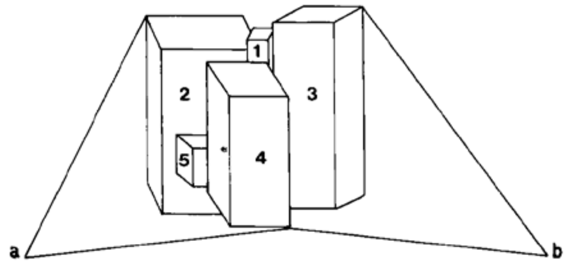


Fig. 2. Two views of a man made database. The boxes are files (relations). a. A plant life form - insect relationship oriented viewer. b. A plant - soil relationships oriented viewer. The reader is a general viewer. 1. Soil parameters; 2. Hemiptera species; 3. Plant species; 4. Plant life forms; 5. Environmental parameters.

parsimonious presentation, in other words, with the least possible number of axes, explaining the highest percentage of trended variation. The second type of reaction although deeply rooted in the first steps of human thought, only recently have been effectively developed (Goodall, 1986), because of the mathematical complexity of the necessary numerical techniques. On the other hand, both ordination and classification are very primitive thought processes and is believed that they occur more or less in many vertebrates and invertebrates (Goodall 1986). For sucking Hemiptera, for instance, these processes are vital besides the fact that they cannot be formally classified as *thought processes*.

Two bottlenecks thus unavoidably exist in the process of information extraction from ecological communities (Fig. 1). The first is created by sampling procedures and the second by the data analysis methods, independently of their inefficiencies and limitations. Between these, lies the *database*, a mere set of data files. *Database system* is the combination of procedural programs and data files, functionally integrated, and *database management system* is the integrated set of programs to support the database, facilitating the manipulations on the level of individual data elements. DAPROPHECO is the acronym of the phrase "data processing in plant Hemiptera communities". *View model* is an aspect of the database seen from a point of the abstract space; a view model is constructed by the database management system (Fig. 2). The definitions of the most important terms are better represented in Fig. 3 and Table 1. The above definitions are not yet generally accepted. Hopefully, the terminology in computer science and especially on database design is chaotic in the face of which ecological terminology may seem to be an arrangement of clearcut items. The overall relation-schema of the database system is presented in the Appendix.

A complete substantiation of the overall relation-schema would take much space and is beyond the scope of this paper. So, I shall try to explain it through a restricted sample of attributes and relations. The reader should consult Table 1 and 2 for the meanings of various symbols and codes.

Example 1. It is supposed that we are aiming to create a file structure to store the collected field data for Hemiptera after an one day survey. We create a relation, and we name it HEMI with the relation-schema represented in Fig. 4. We treat relations as mathematical sets of tuples and this, among other things, requires that the attributes must be ordered and their domains allow only simple values, namely values not decomposable into multiple ones. The single record chosen for presentation does not fulfill these requirements, since in many attributes there are no simple values, e.g. the attribute HEGN contains in its domain the value *Staria, Mycterodus, Dalmatrium* and *Eurydema*. This value can not be decomposed in multiple meaningful values. A solution would be to remove nested groups breaking this record into many meaningful tuples, a process called *normalization into first normal form*. The result can be seen in Fig. 5. In this representation we observe the high degree of redundancy and the difficulties when we attempt to update this file. This HEMI1NORM file is more flexible but its structure is less tightly bound and searching the database for a relationship takes much effort. It is clear that the relation-schema must be broken into smaller subsets because there are functional

relation: **HEMI**

SACO	**DATE**	**TRAN**	**TRPO**	**SIZE**	**VEPH**	[attributes]
16A7	6/22/86	1	12	10	Herb1	t
34W43	5/3/86	2	23	50	Jun-Que	u
2S30	7/30/86	3	125	100	Herb7	p
23R41	2/2/86	1	70	20	Phlomis1	l
1040	12/25/85	3	40	35	Herb7	e
1041	12/25/85.	3	45	10	Phlomis1	s
Domain of SACO	Domain of DATE	Domain of TRAN	Domain of TRPO	Domain of SIZE	Domain of VEPH	

[RELATION-SCHEMA OF rel.:**HEMI**]

Fig. 3. Illustration of the various terms used in relational modelled databases.

relation:HEMI

SACO	DATE	TRAN	TRPO	SIZE	VEPH	HEGN	HESN	HEAU	SUBO	SUPF	FAMI	SUBF	GEDR	MALE	FEMA
16A7	6/22/86	1	2	50	Herb1	Staria	lunata	(Hahn)	Heteropt.	****	Pentatom.	****	wide	2	3
						Mycterodus	pallens	(Stahl)	Auchen.	****	Issidae	****	****	2	0
						Dalmatrium	maculip.	****	Auchen.	****	Issidae	****	Med.	3	4
						Eurydema	ventrale	****	Heteropt.	****	Pentatom.	****	Swide	0	3

INPT	COLE	DIPT	HYME	OTHE	STOR	TIME	TEMP	REHU	LUXO
1,3Dinstar of Staria 2,3Dinstar of Dalmatrium	3	4	2	1	131	09	35	22	350

Fig. 4. Relation HEMI is constructed after an one day survey. For simplicity only one record is illustrated. Explanations; * = in HEAU, SUPF, SUBF denotes that the investigator for unknown reasons is not interested in these values. '2' in TRPO means the distance from one end of the transect in meters. 'Herb 1' in VEPH denotes the vegetation type in which this particular SACO is located. 'wide', 'Med', 'Swide' in GEDR denote a particular pattern of geographic distribution (e.g., 'Swide' means "distributed in South Palearctic regions"). '2,3D instar' in INPT denotes that 2 individuals at third instar stage are collected, belonging to *Dalmatrium maculipes* (Issidae) species. For all other symbols the inspection of Appendix is suggested.**

Table 1. Definitions of terms.

A RELATION is a set of homogeneous tuples. Within it the data elements in a specific raw belong to the same tuple and data elements in a specific column belong to the same attribute within the limits of its specific domain.

A TUPLE is a collection of related heterogeneous atomic data elements partially descriptive of a real world object.

A RELATION-SCHEMA describes the attributes in terms of their domain of values and their participation in the ruling or the dependent part.

The subset of attributes in a tuple which we can consider to name or define the object is the ruling part. The remaining attributes are the dependent part.

NORMALIZATION is the process of removing nested groups.

FIRST NORMAL FORM is the type of resulting relations after the process of normalization.

NORMALIZATION TO SECOND NORMAL FORM in the process of removing functional dependencies on attributes, which are subsets of the ruling part. As a result the dependent part of such a relation, contains only attributes that are functionally dependent on the entire ruling part.

NORMALIZATION TO THIRD NORMAL FORM is the process of removing functional dependencies found among attributes within the dependent part.

In the value of some attributes E is always determined by the value of other attributes A, then we say the E is functionally dependent on A.
formally: $E_1, E_2, \ldots, E_I, \ldots, E_K = FD\,[A_1, \ldots, A_n, \ldots, A_j]$
to avoid redundancy in the ruling part, we require that no subset of the ruling part can be found which would also be an adequate ruling part.

If the value of an attribute determines a set of multiple values then a multivalued dependency occurs.
formally: $MVD\,[A_1, \ldots, A_j] = \{E_{1.1}, \ldots, E_{1.m}\}, \{E_{2.1}, \ldots, E_{2.p}\}$

Table 2. Definitions of operations.

PROJECTION (Π) IS AN OPERATION THAT REDUCES A RELATION $R_i\,(S_i, T_i)$ BY LIMITING ITS ATTRIBUTES.
THE PROJECTION REQUIRES A LIST OF THE NAMES $K_1, \ldots, K_I$
OF THE ATTRIBUTES TO BE EXTRACTED.
formally: $R\,\langle S, T\rangle = \Pi\, R_i\,\langle S_i, T_i\rangle.\ KIK = \{K_1, \ldots, K_I\}$
so that $S = S_i \cap K$

$$t \in T = \pi\, t'' k_1 /\!/ \ldots /\!/ \pi\, t'' k_I\,;\ t'' \in T_i$$

π is an extract operator which obtains a single value ν from a single tuple, ν is the value corresponding to attribute $A_p \in S_i$

SELECTIONS (S) IS AN OPERATION WHICH CREATES A RELATION HAVING THE SAME RELATION SCHEMA BUT CONTAINING A SUBSET OF THE TUPLES, BASED ON THE VALUES OF THEIR ATTRIBUTES.
formally: $R\,\langle S, T\rangle = S\, R_i\,\langle S_i, T_i\rangle / L = \{(A_p\, \Theta\, c_1)\, U \ldots U\, (A_r\, \Theta\, c_I)\}$
so that $t \in T = [(\nu\, t''\, A_p\, \Theta\, c_1)\, U \ldots U\, (\pi\, t''\, A_r\, \Theta\, c_I)]$
with $c_1 \in D\,(S_i, A_p), \ldots, c_I \in D\,(S_i, A_r)$
Θ is one of the comparison operators $=, \langle, \rangle, =\langle, =\rangle$

JOIN (Ξ) IS AN OPERATION WHICH CREATES A NEW RELATION BY COMBINING THE DATA ELEMENTS FROM TWO RELATIONS FORMING THUS A SUBSET OF THE CARTESIAN PRODUCT OF THE TWO RELATIONS. THIS SUBSET CONSISTS OF THOSE TUPLES WHICH HAVE IDENTICAL OR MATCHING VALUES FOR THE JOIN ATTRIBUTE J.
formally: $R\,\langle S, T\rangle = R_1\,\langle S_1, T_1\rangle\, \Xi_{J1\,\Theta\,J2}\, R_2\,\langle S_2, T_2\rangle$
so that $S = S_1 /\!/ S_2$
and each tuple $t \in T = t_1 \in T_1 /\!/ t_2 \in T_2\,;\ \pi\, t_1 . J_1\, \Theta\, \pi\, t_2 . J_2$
when Θ = "=" then we have "equi-join".

dependencies among the attributes defining the object described by the tuple and the attributes which provide the data regarding the object. More precisely the attributes SACO HEGN HESN is the *ruling part* on which all the remaining attributes are functionally dependent, but some of them, say MALE and FEMA are dependent relation:HEMI 1 NORM

SACO	DATE	TRAN	TRPO	SIZE	VEPH	HEGN	HESN	HEAU	SUBO	SUPF	FAMI	SUBF	GEDR
16A7	6/22/86	1	2	50	Herb1	Staria	lunata	(Hahn)	Heteroptera	****	Pentatomidae	****	wide
16A7	6/22/86	1	2	50	Herb1	Mycterodus	pallens	(Stahl)	Auchenorrhyn.	***	Issidae	****	****
16A7	6/22/86	1	2	50	Herb1	Dalmatrium	maculipes	****	Auchenorrhyn.	***	Issidae	****	Med.
16A7	6/22/86	1	2	50	Herb1	Eurydema	ventrale	****	Heteroptera	****	Pentatomidae	****	Swide

MALE	FEMA	INPT	COLE	DIPT	HYME	OTHE	STOR	TIME	TEMP	REHU	LUXO
2	3	1,3Dinstar	3	4	2	1	131	09	35	22	350
2	0	of Staria	3	4	2	1	131	09	35	22	350
3	4	1,3Dinstar	3	4	2	1	131	09	35	22	350
0	3	of Dalmatr.	3	4	2	1	131	09	35	22	350

Fig. 5. The relation HEMI1NORM is the first normalized form of relation HEMI (explanations in the text).

on a subset of this part, namely HEGN and HESN. It is evident that a procedure is needed to decompose this schema into more compact, easily updatable, though loosely bound schemas. More formally, the relation HEMI1NORM must be further normalized to remove dependencies on subsets of the ruling part, that is to render it into its *second normal form* and to remove dependencies among attributes within the dependent part, that is to render it into its *third normal form*. The result is a set of separate relations and the relationship among them is expressed through common domains as illustrated in Fig. 6. In this overall schema, the decomposition has proceded beyond mathematical requirements; in relations HEM2 and HEM4 the ruling part SACO has entirely different, not functionally related, dependent parts, namely DATE, TRAN, TRPO, SIZE, VEPH and TIME, TEMP, REHU, LUXO respectively. However, this decomposition is dictated for practical reasons; for each sample the first set of attributes as a rule is always recorded, but this is not true for the second set. If we had updated these files into one, we should have the problem of several null entries in the joined file in addition to the time consuming input operations.

Retrival and relation algebra

The final structure of the database which consists of normalized files, reduces redundancy, speeds up searching processes and saves storage space, but requires an effective algebraic thought on a set-of-files level. In relational modelled databases several operations on files are available. In taxonomy and systematics the requirements for a management system manipulating a set of files was rarely a vital demand. In this disciplines a monofiled database is usually the appropriate data organization type and network and hierarchical models

relation:HEM0

HESC	HEGN	HESN	HEAU	SUBO	SUPF	FAMI	SUBF	GEDR
Stalun	Staria	lunata	(Hahn)	Heteroptera	****	Pentatomidae	****	wide
Mycpal	Mycterodus	pallens	(Stahl)	Auchenorrhyncha	****	Issidae	****	****
Dalmac	Dalmatrium	maculipes	****	Auchenorrhyncha	****	Issidae	****	Med.
Eurven	Eurydema	ventrale	****	Heteroptera	****	Pentatomidae	****	Swide

relation:HEM1

SACO	HESC	MALE	FEMA	INPT
16A7	Stalun	2	3	1,3Dinstar of Staria
16A7	Mycpal	2	0	1,3Dinstar of Dalmatrium
16A7	Dalmac	3	4	
16A7	Eurven	0	3	

relation:HEM2

SACO	DATE	TRAN	TRPO	SIZE	VEPH
16A7	6/22/86	1	2	50	Herb1

relation:HEM3

SACO	COLE	DIPT	HYME	OTHE	STOR
16A7	3	4	2	1	131

relation:HEM4

SACO	TIME	TEMP	REHU	LUXO
16A7	09	35	22	350

Fig. 6. The five relations (HEMO, HEM1, HEM2, HEM3, HEM4) are the result of the decomposition of the relation HEMI1NORM after the second and third normalization (see text).

relation:LEXE

TRAN	TRPO	SIZE	SLOP	EXPO	MASL	QUCO
1	0	2	0.100	45	70	A20
1	2	2	0.100	45	70	A21
1	4	2	0.150	45	68	A22
1	6	2	0.090	45	68	A23
1	8	2	0.100	45	68	A24
1	10	2	0.150	45	68	A25
1	12	2	0.150	45	67	A26
1	14	2	0.150	45	67	A27
1	16	2	0.150	45	67	A28
1	18	2	0.100	45	66	A29
1	20	2	0.100	45	66	A30
2	0	2	0.200	00	02	B62
3	0	2	0.250	180	80	B72

relation:REL2

TRAN	TRPO	QUCO
1	0	A20
1	2	A21
1	4	A22
1	6	A23
1	8	A24
1	10	A25
1	12	A26
1	14	A27
1	16	A28
1	18	A29
1	20	A30

```
REL1= Π (LEXE).K|K={TRAN,TRPO,QUCO}

REL2= S (REL1)|L={TRAN=<1}
```

Fig. 7. Relation LEXE is projected on the attributes {TRAN, TRPO, QUCO} and the result relation REL1 (not illustrated) is selected for those tuples meeting the condition TRAN = ⟨1. The final relation is REL2.

were the adequate data management solutions (Wildi, 1986). But in community studies information is always located in different files the number of which is always increased as a result of normalization processes.

On a set-of-files level seven operations are available, four of which are the conventional set operations: union, intersection, set difference and cartesian product. These are of minor importance in DAPROPHECO. The remaining three are of special importance. In Table 2 we can see the verbal and formal definitions of *projection* Π *selection* S and *join* Ξ operations. Their detailed usage will be illustrated in a sample run with simplified data sets. With real data sets the basics of the procedure are identical but the volume of the output and the risk of overflows is usually a serious factor which must be taken always in mind. The job has been done on an IBM compatible processor with 640 kb RAM and 80 Mb hard disk memmory capacity; additionally 18 files are allowed to be simultaneously opened. The procedural language used, that is the language which let us get a solution is the dBase III. This is a higher level language than a programming language such as FORTRAN 77, Pascal or Basic, although not fully automated in relational algebra statements such as the IBM System-R and its ancestor SQL/DS. Additionally it has the ability to create files for programming languages such as Basic and FORTRAN 77 through programs written by the database designer. This facilitates arrangements of data sets for further processing and analysis through multivariate statistical methods and graphic representation. Before any retrieval task the questions asked by the system must be clearly defined in the context of the relation schemas. The user is presented with warning signals to avoid abortions of execution or overflows.

Example 2. Let us suppose that we suspect that the essential oils in the leaves of some plant species account for the occurence of some Hemiptera species on a specific transect. Also we anticipate that the average essential oil content of the leaves of a specific plant species provides not enough information for such analyses and we desire some additional parameters of performance in various transect positions such as coverage, density and mean height. But three parameters for plant performance are too many to serve inspection purposes and inadequate for further multivariate analysis of this type; so we decide to adopt the three factor summed dominance ratio (SDR3) of Numata (1979), that is the average of the 3 parameters for each plant species. We believe that this procedure will give us important information on the quantity of essential oil content for each plant species on a transect position basis. Also, we assume that we are not interested in variations of our parameters in time.

The relevant database files are PLAD7, PLADOM2, HEM2, LEXA, LEXB, LEXC, LEXD and LEXE. We begin by selecting from LEXE the QUCOs for TRAN = 1 (Fig. 7). This operation assumes the searching of the entire file LEXE. So we decide to initiate the job by projecting the LEXE file on the attributes TRAN TRPO and QUCO; immediately after this we select those tuples matching the value 1 for the attribute TRAN. These operations are formally written at the end of the file listing. Also, we may notice, at first glance some redundancy. However this is illusive; the attributes SIZE and EXPO are not unique throughout an individual transect. Intermediate files created during the job must be named and the user is presented by a list of options regarding their storage or deletion. In the same figure - right half - the refult relation is represented. The next step is the joining of relations REL2 and PLADOM2, with the difference that DATEs and the attribute SOCIability are no longer of interest and we desire the SDR3 along with the COVErage DENSity and MeanHEIght. What we need is a projection of the relation PLADOM2 on attributes QUCO, PLSC, COVE, DENS, MHEI and the creation of an additional not previously defined domain, that is SDR3. In Fig. 8 we can see the resultant relation REL3, under the operations applied for its creation. The necessary work to complete this section of our job is the adjoining of average leaf essential oil content to each plant species. We focus our interest on relations LEXB, LEXC, LEXD and PLAD7 from which we select those tuples matching the PLSCs already occuring in relation REL3. The procedure is expressed in Fig. 9 and the resulting relation follows under the name RELB; the attributes INDC, BRCO, BCTC, are removed by projection. The LeafEssentialOilContent is expressed as the percentage of dry weight loss after essential oils extracrelation:REL3

PLSC	COVE	DENS	MHEI	SDR3	TRPO
Pister	5	1	27	15.0	0
Balace	2	2	7	3.6	0
Pramaj	3	2	8	4.3	0
Phlfru	3	1	11	5.0	2
Alymin	0.5	15	0.8	5.4	4
Dasvil	2	40	7	16.3	4
Urgmar	0.5	2	1	1.2	6
Phlfru	6	2	12	6.7	6
Dasvil	3	50	8	20.3	8
Hyphir	1	1	12	4.7	8
Urgmar	0.5	3	1	1.5	8
Cycgra	0.5	8	0.5	3.0	10
Phlfru	4	2	7	4.3	10
Brohor	1	100	1	34.0	10
Pislen	9	3	8	6.7	12
Pramaj	1	2	4	2.3	12
Junpho	10	2	30	14.0	14
Allsub	8	60	2	23.3	16
Pislen	7	1	1	3.0	18
Dasvil	1	20	6	9.0	18
Quecoc	5	1	12	6.0	20
Allsub	1	13	3	5.7	20
Urgmar	2	7	2	3.7	20

REL3= Π (PLADOM).K|K={TRPO, PLSC,COVE,DENS,MHEI,SDR3} where SDR3=average [COVE,DENS,MHEI]

Fig. 8. The database relation PLADOM is projected on the attributes {TRPO, PLSC, COVE, DENS, MHEI, SDR3}. Attribute SDR3 is created by the user with values equal to (COVE + DENS + MHEI)/3. Only the resulting relation is illustrated.

relation:RELB

PLSC	COVE	DENS	MHEI	SDR3	TRPO	LEOA
Pister	5	1	27	15.0	0	12
Balace	2	2	7	3.6	0	46/52
Pramaj	3	2	8	4.3	0	22
Phlfru	3	1	11	5.0	2	40
Alymin	0.5	15	0.8	5.4	4	****
Dasvil	2	40	7	16.3	4	****
Urgmar	0.5	2	1	1.2	6	****
Phlfru	6	2	12	6.7	6	40
Dasvil	3	50	8	20.3	8	****
Hyphir	1	1	12	4.7	8	****
Urgmar	0.5	3	1	1.5	8	****
Cycgra	0.5	8	0.5	3.0	10	****
Phlfru	4	2	7	4.3	10	40
Brohor	1	100	1	34.0	10	****
Pislen	9	3	8	6.7	12	15
Pramaj	1	2	4	2.3	12	22
Junpho	10	2	30	14.0	14	10/6
Allsub	8	60	2	23.3	16	****
Pislen	7	1	1	3.0	18	15
Dasvil	1	20	6	9.0	18	****
Quecoc	5	1	12	6.0	20	12/6
Allsub	1	13	3	5.7	20	****
Urgmar	2	7	2	3.7	20	****

REL B = (REL 3) Ξ_{PLSC} Π[(LEXB) Ξ_{INDC} (LEXC) Ξ_{BRCO} (LEXD) Ξ_{BCTC} Π(PLAD7).K_1 |K_1={BCTC , LEOA}

Fig. 9. The procedures which resulted in the creation of RELB. See the text for explanations.

tion by the method of Hughes (1970). In relation RELB, some tuples differ only in the values of the attribute LEOA, because there are more than one analyses of essential oil yield. In plants adapted to mediterranean climates, there is considerable non trended variation in the essential oil content (Morrow and Fox, 1980) and the removal of one value or the substitution by an average is not generally recomended. It is better to leave the structure of this relation as it is. Also, some tuples in the domain LEOA are lacking; for various reasons essential oils are not extracted (e.g. *Urginea maritima*), or as is the case of grasses, the values of dry weight loss after the extraction are less than 2%. We are confronted with a problem here rooting in the project design but this can be handled by putting arbitrarily for all grasses zero values and for non grasses the critical value 2%.

The same procedures hold for the work on the Hemiptera section. The relevant relation is tabulated in Fig. 10 and the operations for its generation are written above. The domain of attribute SIZE contains the number of sweep nettings applied in the respective transect position, in which sampling of the vegetation has been done. Also the attribute VEgetationPHase is not an instantaneous assessment of the vegetation status at the momment of sampling. It is the result of previous cluster analyses in which vegetation data have been submitted.

In the previous process we notice at first glance some waste of effort related to the maintenance of ma-

relation:HEMIPTERA

TRAN	TRPO	SACO	SIZE	HESC	MALE	FEMA	VEPH
1	0	30	10	Dalmac	1	2	Lab1
1	0	30	10	Gargen	0	1	Lab1
1	0	30	10	Stalun	1	1	Lab1
1	4	31	15	Rannew	1	0	Herb1
1	4	31	15	Mycpal	1	0	Herb1
1	10	32	20	Dalmac	1	1	Herb/rock
1	10	32	20	Quabas	0	2	Herb/rock
1	10	32	20	Melmat	5	4	Herb/rock
1	10	32	20	Stalun	2	3	Herb/rock
1	15	33	10	Mycpal	2	1	Junpho
1	20	34	30	Quabas	4	1	Pist_Quer
1	20	34	30	Melmat	4	0	Pist_Quer
1	20	34	30	Mycpal	3	4	Pist_Quer

(HEMIPTERA)={ S [Π (HEM1).K|K={SACO,TRAN,TRPO,SIZE}] .L|L={(TRAN=<1)}} Ξ_{SACO} [Π (HEM2).K1|K1={SACO, HESC,MALE,FEMA}]

Fig. 10. Explanations in the text.

a

180	0	0	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0
0	200	0	0	0	172	0	0	0	0	0
0	0	11	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	2	3	0	0	0	0	0	8
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	6	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	100	0	0	45	0
95	0	0	0	0	0	51	0	0	0	0
0	0	0	0	0	0	0	140	0	0	0
0	0	0	0	0	0	0	0	47	0	12
0	0	0	0	0	0	0	0	0	0	72

b

3	0	2	0	0
1	0	0	0	0
2	0	5	0	0
0	1	0	0	0
0	1	0	3	7
0	0	2	0	5
0	0	9	0	4

Fig. 11. Two BASIC data files readable by the Orloci and Kenkel package (see text for explanations). Columns are TRPOs and rows are plant species (a) and Hemiptera species (b).

nagerial files LEXA, LEXB, LEXC and LEXD. In the overall database structure I have made provision for several views, provided there are no conflicts among them. It is a specialized database for integrated studies and the file serve many purposes. For example the files LEXA, LEXB, LEXC, LEXD, LEVE and PLAD7 are relevant in investigation on dimension analysis of the plants as well as in pure physiological studies.

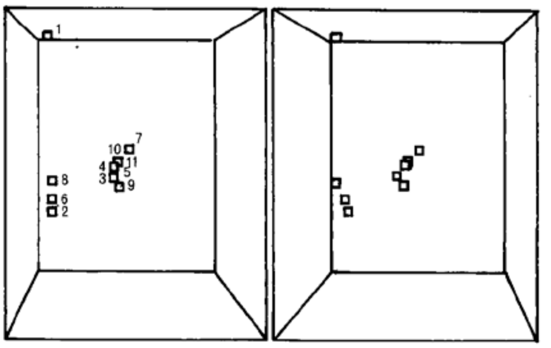
We have finished our job and we are presented with a peculiar situation. Data on plants are more systematically taken from the transect and the relative sampling units are of the same size. But in the relation HEMIPTERA, the sampling units are of various sizes and are located at different distances. In addition, the sizes of sampling units in these two relations are not directly comparable. A further analysis is necessary and what we need is a strategy to pass through the next bottleneck. As is usually the case in community studies, we subjectively select an objective method. The only constraint is the justification of our decision.

DAPROPHECO management system has embodied procedures to copy the contents of selected domains into data files readily accessed by programming languages. In Fig. 11 we can see two BASIC text files ready for further processing through the Orloci and Kenkel (1985) package. Columns are transect positions and rows are species scores. Plant species score is the product of SDR3 and LEOA and Hemiptera species score is the sum of MALEs and FEMAles. Without full justification, Fig. 12 and 13 present graphically the outputs of an ordination technique - multidimensional scaling - applied to both files. In these stereograms, TRansect POsitions are ordinated on three species axes. Also a program is in progress to interface the DAPROPHECO system with the Wildi and Orlòci (1982) package.

Additionally, an agglomerative clustering technique, Orloci's sum of squares agglomeration, is applied on the same files and the TRPOs are considered as individuals. The respective dendrograms follow in Fig. 14 and 15. However, it is meaningless to discuss the output on an ecological basis, because the data are an oversimplified reduction of real data. Tuples represent real records, but their allotment into one relation is an artifact, serving illustrative purposes.

Conclusion

The relational model for databases is of special importance in the study of ecological communities. Their importance in phytosociology has been emphasized only recently (see Wildi, 1986). DAPROPHECO is a realization of the most recent developments in the discipline of database design. It is a software package written in the procedural language dBase III, implemented under MS operating system on an Advanced Technology Systems machine. The package has not taken its final



b

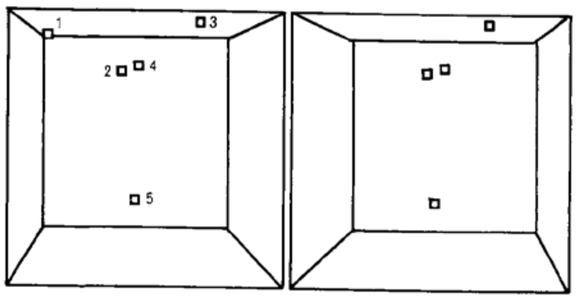
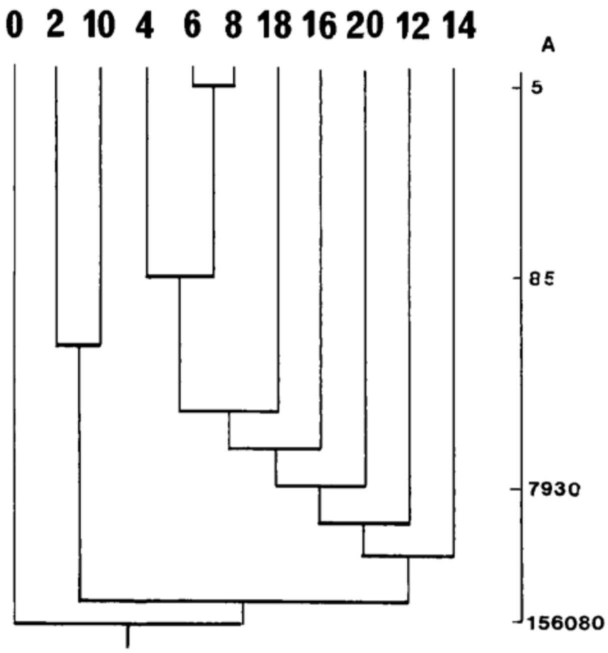


Fig. 12-13. Output of the Orloci and Kenkel package (program STEREO, see text). This type of graphics are stored as graphic files by the system, reducing thus the waste of time very often met in graphic jobs with BASIC, a. Stereogram of TRPOs based on plant species performances (output from a multidimensional scaling algorithm, program MDSCAL). b. Stereogram of TRPOs based on Hemiptera species abundances (with the same algorithm).



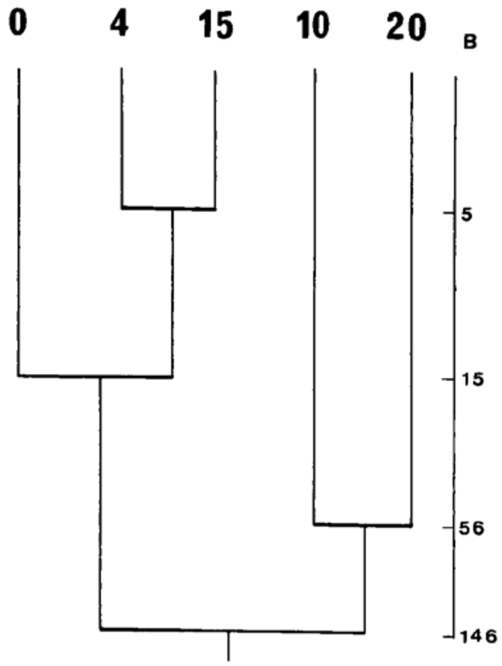


Fig. 14-15. Cluster analysis output of the Orloci and Kenkel package (dissimilarity values are in logarithmic scale, program TREE; see text for more explanations). Dendrograms are stored in graphic files by DAPROPHECO. a. TRPOs clusters based on plant performances. b. TRPOs clusters based on Hemiptera abundances.

form, but I hope that this will happen in the near future. The structure of the database can be changed at will. These changes of course assume respective changes or further developments of the programs in the data management system which is a disadvantage of the system. To cope with these problems, I decided to modify the system on a menu-driven fashion. This not only facilitates the unexperienced user but enables him to formalize better the problems and the developmental necessities of the system. The user is presented with a table of options and he selects the files and the jobs he intends to do with the system.

Aknowledgements. This work is part of a project designed for the community level analysis of plant/Hemiptera associations in east-Mediterranean ecosystems. Support was received by BORA Software House and the discussions with all the good people there, are greatly aknowledged. Thanks are extended to K. Syrakos and E. Katsis for all helpful and good humoured discussions and advice about IBM and ATS machines. Also, I thank prof. P.J. van Albada, Dr. S. Drosopoulos, Dr. P. Ganis, prof. E. Feoli and prof. L. Orlóci for comments on the manuscript and suggestions.

REFERENCES

Astrahan, M.M., M.W. Blasgen, D.D. Chamberlin, K.P. Eswaran, J.N. Gray, P.P. Griffiths, W.F. King, R.A. Lorie, P.R. McJohnes, J.W. Mehl, G.R. Putolu, I.L. Traiger, B.W. Wade and V. Watson. 1976. System R: Relational approach to database management. TODS 1: 97-137.

Braun-Blanquet, J. 1964. *Pflanzensoziologie*. 3rd ed. Springer-Verlag. Wien.

Castro, E.F., J. Hanson, T. Rettig. 1985. *Advanced programmers guide: featuring dBase III and dBase II*. Ashton-Tate. California.

Cood, E.F. 1970. A relational model of data for large shared data banks. CACM 13: 377-387.

Cood, E.F. 1982. Relational database: a practical foundation for productivity. CACM 25: 109-117.

Estabrook, G.F. and R.C. Brill. 1969. The theory of the TAXIR accessioner. Math Bios 5: 327-340.

Goodall, D.W. 1986. Classification and Ordination: their nature and role in taxonomy and community studies. Coenoses 1: 3-9.

Greig-Smith, P. 1983. *Quantitative Plant Ecology*. 3rd ed. Blackwell, Oxford.

Hughes, A. 1970. A modified receiver for heavier than water essential oils. Chem. Ind. (London) 48: 1536.

Morrow, P.A. and L.R. Fox. 1980. Effects of variation in *Eucalyptus* essential oil yield on insect growth and grazing damage. Oecologia 45: 209-219.

Numata, M. 1979. The structure and succession of grassland vegetation. In: M. Numata (ed), *Ecology of Grasslands and Bamboolands in the World*, p.p. 182-196. Gustav Fischer Verlag, Jena.

Orloci, L. 1978. *Multivariate Analysis in Vegetation Research*. Jung, The Hague.

Orloci, L. and N.C. Kenkel. 1985. *Introduction to Data Analysis with Applications in Population and Community Biology*. ICPH, Maryland, USA.

Pielou, E.C. 1977. *Mathematical Ecology*. Wiley, New York.

Wildi, O. 1986. The relational model for databases in community studies. Coenoses 1: 29-34.

Wildi, O. and L. Orloci. 1982. *Management and Multivariate Analysis of Vegetation Data*. 2d ed. SFIRF, Swiss.

APPENDIX - Overall relations

DA.PRO.P.HE.CO.

C = CHARACTER DATA TYPE
N = NUMERICAL DATA TYPE
D = CALENDAR DATA TYPE
M = MEMO DATA TYPE NO FURTHER PROCESSED
L = LOGICAL DATA TYPE (TRUE OR FALSE)
THE NUMBER AFTER THE DATA TYPE CHARACTERIZATION DEFINES THE LENGTH OF THE FIELD AND THE NUMBER OF DECIMAL PLACES.

BLOCK:HEMIPTERA

RELATION:HEM0

HEGN	C	15	GENUS NAME
HESN	C	15	SPECIES NAME
HEAU	C	15	AUTHORITY NAME
SUBO	C	4	SUBORDER NAME
SUPF	C	4	SUPERFAMILY NAME
FAMI	C	15	FAMILY NAME
SUBF	C	4	SUBFAMILY NAME
HESC	C	6	HEMIPTERAN SPECIES CODE
GEDI	C	10	GEOGRAPHICAL DISTRIBUTION

RELATION:HEM2

TRAN	C	3	TRANSECT CODE
TRPO	C	4	TRANSECT POSITION
SIZE	N	2/0	SIZE OF THE SAMPLE EXPRESSED AS NUMBER OF SWEEP NETTINGS OR VOLUME OF AIR PASSED THROUGH A VACUUM NET
VEPH	C	6	VEGETATION PHASE DETERMINED BY OBJECTIVE ANALYSIS OF VEGETATION DATA
DATE	D	8	DATE
SACO	C	4	CODE OF THE SAMPLE TAKEN

RELATION:HEM1

SACO	C	4	****
HESC	C	6	****
MALE	N	4	NUMBER OF MALE INDIVIDUALS IN THE SPECIFIED BY SACO SAMPLE
FEMA	N	4	NUMBER OF FEMALE INDIVIDUAL IN THE SPECIFIED BY SACO SAMPLE
INPT	M	2	NUMBER OF LARVAE IN THE SPECIFIED BY SACO SAMPLE

RELATION:HEM3

SACO	C	4	****
COLE	N	2	NUMBER OF COLEOPTERA SPECIES COLLECTED
DIPT	N	2	NUMBER OF DIPTERA SPECIES COLLECTED
HYME	N	2	NUMBER OF HYMENOPTERA SPECIES COLLECTED
OTHE	N	2	NUMBER OF OTHER SPECIES COLLECTED
STOR	C	3	LABEL OF THE SACO FOR STORAGE

RELATION:HEM4

SACO	C	4	****
TIME	N	2/0	TIME OF THE DAY
TEMP	N	3/1	TEMPERATURE
REHU	N	2/0	RELATIVE HUMIDITY
LUXO	N	3/0	LUXOMETRIC MEASUREMENT

BLOCK:PLANT DOMINANCE

```
----------------------------------
RELATION:LEXF

GNAM C 15     GENUS NAME
SNAM C 15     SPECIES NAME
AUTH C 15     AUTHORITY NAME
FNAM C 15     FAMILY NAME
GEDI C 10     GEOGRAPHIC DISTRIBUTION
PLSC C  6     PLANT SPECIES CODE


----------------------------------
RELATION:LEXE

TRAN C   3    TRANSECT OR PLOT ON WHICH THE QUCO OCCURS
TRPO C   4    TRANSECT POSITION OF THE RESPECTIVE QUCO
SIZE N 4/0      SIZE OF THE QUCO IN SQUARED DECIMETERS
SLOP N 4/3      SLOPE OF THE QUCO
EXPO N 3/0      EXPOSURE OF THE QUCO IN DEGREES
MASL N 4/0      METERS ABOVE SEA LEVEL
QUCO C   4    QUADRAT CODE OR MORE GENERALLY SAMPLING UNIT CODE


----------------------------------
RELATION:LEXA

QUCO C   4              ****
PLSC C   6              ****


----------------------------------
RELATION:LEXB

PLSC C   6              ****
INDC C   4    CODE OF PLANT INDIVIDUAL IN CASE OF TREES SHRUBS OR TUFTED PERENNIAL HERBS


----------------------------------
----------------------------------
RELATION:PLADOM1FOR TREES SHRUBS AND TUFTED PERENNIALS

QUCO C   4              ****
PLSC C   6              ****
INDC C   4              ****


----------------------------------
RELATION:PLADOM1A

INDC C   4              ****
HEIG N 4/0      HEIGHT IN DECIMETERS
DBHE N 4/0      DIAMETER AT BREAST HEIGHT IN MILLIMETERS
BAAR N 4/0      BASAL AREA IN SQUARED CENTIMETERS
CRAD N 3/0      CROWN AVERAGE DIAMETER
DATE D   8    DATE


----------------------------------
RELATION:PLADOMFOR ALL PLANTS
QUCO C   4          ****
PLSC C   6          ****
COVE N 3/0     COVERAGE IN SQUARED DECIMETERS
DENS N 3/0     NUMBER OF PLANT INDIVIDUALS OCCURRING ON QUCO
MHEI N 4/1     MEAN HEIGHT OF PLANTS ESTIMATED BY EYE IN DECIMETERS
SOCI C   1    SOCIABILITY INDEX SENSU BRAUN-BLANQUET EXTENDED
DATE D   8    DATE
```

RELATION:PLADOM3 FOR REGRESSIONS ON HERBS

```
INDC C  4        ****
STHE N  3/0   STEM HEIGHT IN CENTIMETERS
STBD N  4/2   STEM BASAL DIAMETER IN MILLIMETERS
STDW N  6/3   STEM DRY WEIGHT IN MILLIGRAMMS
ROLE N  3/0   ROOT LENGTH IN MILLIMETERS
RODW N  6/3   ROOT DRY WEIGHT IN MILLIGRAMMS
ROIP N  2/0   ROOT INTERSECTION POINTS ON A GRID OF LINES 5 MILLIMETERS APART
LBNU N  3/0   NUMBER OF GREEN BASAL LEAVES
LSNU N  3/0   NUMBER OF GREEN STEM LEAVES
FRNU N  3/0   NUMBER OF FRUITS ON THE WHOLE PLANT
FLNU N  3/0   NUMBER OF FLOWERS OR INFLORESCENCES IN CASE OF MINUTE FLOWERS
                         AS IN GRAMINAE AND EUPFORBIAS.
```

BLOCK:**PLANT DIMENSION ANALYSIS**

RELATION:LEXC

```
INDC C  4        ****
BRCO C  5     BRANCH CODE
```

RELATION:LEXD

```
BRCO C  5        ****
BCTC C  6     CURRENT OR PREVIOUS YEAR TWIG CODE
```

RELATION:PLAD2

```
INDC C  4        ****
DATE D  8     DATE
STBD N  4/0   STEM BASAL DIAMETER IN MILLIMETERS
STHE N  4/0   STEM HEIGHT IN DECIMETERS
STAG N  3/0   STEM AGE IN YEARS
STGO M  10    MEMO FIELD CONTAINING GENERAL OBSERVATIONS
BRNU N  2/0   NUMBER OF BRANCHES ARISING DIRECTLY FROM THE MAIN STEM
BRCO C  5        ****
```

RELATION:PLAD3

```
BRCO C  5        ****
BRDT N  4/0   BRANCH DISTANCE FROM THE TIP OF THE MAIN STEM
BRBD N  5/2   BRANCH BASAL DIAMETER IN MILLIMETERS
BRAG N  2/0   BRANCH AGE IN YEARS
BBTH N  4/2   BRANCH BARK THIKNESS
BBDW N  6/3   BRANCH BARK DRY WEIGHT
BWDW N  6/3   BRANCH WOOD DRY WEIGHT
BRLE N  3/0   BRANCH LENGTH IN CENTIMETERS
BRNG N  2/0   BRANCH NUMBER OF GENERATIONS
BRTF N  4/2   BRANCH TAPER FACTOR
BNDP N  2/0   BRANCH NUMBER OF DEAD PARTS E.G. BRANCHLETS, TWIGS
BTDW N  6/3   BRANCH TRUE DRY WEIGHT
BCTN N  3/0   BRANCH CURRENT TWIGS NUMBER
BCTW N  6/3   BRANCH CURRENT TWIGS WEIGHT
```

RELATION:PLAD4

```
BCTC C  5        ****
CTLN N  2/0   NUMBER OF LEAVES
CTMD N  1/0   MEAN DAMAGE EXPRESSED IN AN ARBITRARY SCALE 0-10
CTNI N  2/0   NUMBER OF INFLORESCENCES OR SINGLE FLOWERS
CTNF N  2/0   NUMBER OF FRUITS
```

RELATION:LEAF

```
BCTC  C   5        ****
LECO  C   3   LEAF CODE
```

RELATION:PLAD5

```
LECO  C   5        ****
LBLE  N  5/2  LEAF BLADE LENGTH IN MILLIMETERS
LBWI  N  4/2  LEAF BLADE WIDTH IN MILLIMETERS
LPLE  N  4/2   LEAF PETIOLE LENGTH IN MILLIMETERS
LDWU  N  7/4  LEAF DRY WEIGHT PER UNIT AREA IN 100MICROGRAMMS
LDWE  N  7/4  LEAF DRY WEIGHT IN 100MICROGRAMMS
LDTH  N  4/0  LEAF DRY THIKNESS IN MICRONS
LPER  N  2/0  LEAF PERSISTANCE IN MONTHS
```

RELATION:PLAD6CHLOROPHYLL CONTENT

```
BCTC  C   5        ****
LCHC  N  3/1    CURRENT LEAF CHLOROPHYLL CONTENT IN MICROGRAMMS PER SQUARED CENTIMETER OF BLADE AREA
LCHS  N  3/1    SECOND YEAR LEAF CHLOROPHYLL CONTENT IN UNITS AS ABOVE
LCHA  N  3/1    AVERAGE LEAF CHLOROPHYLL CONTENT IN UNITS AS ABOVE
```

RELATION:PLAD 7ESSENTIAL OIL YIELD

```
BCTC  C   5        ****
LEOC  N  3/1  CURRENT LEAF ESSENTIAL OIL YIELD IN MILLIGRAMMS PER GRAM OF DRY LEAF WEIGHT
LEOS  N  3/1  SECOND YEAR LEAF ESSENTIAL OIL YIELD
LEOA  N  3/1  AVERAGE LEAF ESSENTIAL OIL YIELD
```

RELATION:PLAD8NUTRIENTS AND WATER OF LEAVES

```
BCTC  C   5        ****
LENI  N  4/3  PERCENTAGE OF N CONTENT
LEMG  N  4/3  PERCENTAGE OF Mg CONTENT
LEPH  N  4/3  PERCENTAGE OF P CONTENT
LEWA  N  3/1  PERCENTAGE OF WATER CONTENT
```

RELATION:PLAD9ROOT SYSTEM

```
INDC  C   4        ****
RTDW  N  7/4  ROOT DRYWEIGHT IN KILOGRAMS
RTFA  N  4/2  ROOT TAPER FACTOR
RLEN  N  3/0  ROOT LENGTH IN CENTIMETERS
RWID  N  3/0  ROOT WIDTH IN CENTIMETERS
```

BLOCK:PHENOLOGY

RELATION:BRADDRBRANCHES OF WHICH THE RESPONSES ARE MONITORED

```
BRCO  C   5        ****
BRHE  N  4/0  BRANCH HEIGHT IN CENTIMETERS
BREX  N  3/0  BRANCH EXPOSURE IN DEGREES
STEH  N  3/0  STEM HEIGHT IN DECIMETERS
STBD  N  4/0   STEM BASAL DIAMETER IN MILLIMETERS
SOIL  C  20   SOIL TYPE ON WHICH THE BRANCH AND NOT THE WHOLE PLANT OCCURS
```

```
----------------------------------
RELATION:BRRESP BRANCH RESPONSES

BRCO C   5          ****
DATE D   8     DATE
BRBD N 4/2     BASAL DIAMETER IN MILLIMETERS
BRLE N 3/0     LENGTH IN MILLIMETERS
STCA N 6/3     STEM CAMBIAL ACTIVITY IN MILLIMETERS
FLLA N 2/0     FLOWER BUDS ACTIVATED
FLEA N 2/0     FLOWERS WITH COROLA PRESENT BUT NOT FULLY DEVELOPED
FLGA N 2/0     FLOWERS WITH FULLY DEVELOPED COROLA
FLCB N 2/0     FLOWERS WITH THE FIRST SIGNS OF POLLINATION
FLCC N 2/0     FLOWERS WITH PETALS OR PERIANTH SEGMENTS MORE OR LESS FALLEN
FRCD N 2/0     FRUITS GREEN PRESENT
FRCE N 2/0     FRUITS MATURE OR RIPEN
LELF N 2/0     LEAF BUDS ACTIVATED
LEEF N 2/0     LEAVES SMALL OR VISIBLE ONLY AS GREEN SPOTS
LEGF N 2/0     LEAVES SMALL BUT WELL DEVELOPED THOUGH TENDER AND LIGHT GREEN
LECG N 2/0     LEAVES FULLY DEVELOPED
LECH N 2/0     LEAVES WITH THE FIRST SIGNS OF DECAY BUT NOT YET FALLEN
BRIN N 2/0     BRANCH INFECTIONS AS PERCENTAGE OF THE SURFACE
LMDB N 2/0     LEAF MEAN DAMAGE FROM MARGINAL BITES EXPRESSED AS PERCENTAGE OF THE TOTAL BLADE AREA
LMDS N 2/0     LEAF MEAN DAMAGE EXPRESSED AS AREA OF BLADE AREA COVERED WITH YELLOW BROWN OR
           TRANSPARENT SPOTS.

----------------------------------
```