

---

## IV – APLICACIONES DESENVOLUPADES

---



*Liberation*, 1955

M.C. Escher



## Capítol 4.1: *Araucaria*, un sistema de determinació automàtica d'inventaris

### 4.1.1 Introducció: Sistemes basats en el coneixement

Un sistema basat en el coneixement (*knowledge-based system, KBS*) o sistema expert (*expert system*) és un programa informàtic que emula, en un determinat camp, el procés que realitza un expert en la solució d'un problema (Xu *et al.* 2001). Un sistema basat en el coneixement consta d'un cos o base de coneixement (*knowledge base*) i un motor d'inferència (*inference engine*). La base de coneixement emmagatzema regles i fets derivats del coneixement expert en una base de dades, de manera anàloga al cervell humà. D'altra banda, el motor d'inferència realitza deduccions lògiques i síntesi de coneixements, per tal de generar solucions per a un problema específic.

Segons Watermann (1986) un sistema expert té dues maneres complementaries de relacionar-se amb l'usuari: el mode d'adquisició de coneixement i el mode de consulta. El mode d'adquisició es relaciona amb usuaris experts per a modificar les regles i els fets del cos de coneixement. El mode de consulta permet respondre les preguntes d'usuaris no experts. Aquesta particular capacitat del sistema expert de modificar dinàmicament les seves regles de decisió fa que hom pugui atribuir-li una certa capacitat d'aprenentatge.

A la literatura es troben alguns treballs sobre el desenvolupament de sistemes experts aplicats a la vegetació. Noble (1987) proposà la definició de sistemes experts per a donar respostes expertes en qüestions d'ecologia aplicada, sobretot a la gestió del territori o la gestió de desastres ecològics. Rykiel (1989) estudià l'ús d'aquests sistemes en la gestió de recursos naturals. Aquesta línia de treball ha donat fruits, per exemple, per la gestió de comunitats pratenses (Xu *et al.* 2001).

Els treballs de Camiz (1988, 1993) s'orienten a promoure un sistema expert per a l'anàlisi de comunitats, mitjançant tècniques multivariants. Segons Camiz (1988), un sistema expert en vegetació "*should be able to assign a particular relevé to a syntaxonomic unit, according to the presence or absence of the leading species, check for the ecological characters of the synecological unit, suggest peculiarities in the vegetation pattern, etc.*". No hem sabut trobar, a part dels treballs teòrics de Camiz, referències bibliogràfiques relacionades amb el desenvolupament de sistemes experts per a la determinació automàtica d'inventaris de vegetació.

## 4.1.2 Definició del sistema de determinació

### 4.1.1.1 Objectius

L'objectiu d'aquesta secció és descriure el desenvolupament i funcionament d'una eina de *software* (programari *Araucaria*, <http://biodiver.bio.ub.es/vegana/araucaria>) que gestiona el coneixement d'un seguit d'unitats de vegetació. El conjunt d'inventaris i les unitats de vegetació establertes constitueixen el 'cos de coneixement' de l'eina. Mitjançant tècniques de classificació numèrica es pretén donar resposta a la consulta d'usuaris no experts en relació a la determinació de nous inventaris de vegetació. Anomenarem aquesta eina 'sistema de determinació automàtica', evitant la denominació de sistema expert, denominació que considerem de moment massa ambiciosa respecte al programari desenvolupat fins ara.

És important diferenciar, d'una banda un sistema expert merament com a eina d'interacció entre l'expert en un camp i un usuari no expert, i de l'altra un sistema que permeti modificar el mateix coneixement de l'expert. En el primer, el sistema es basa solament en un seguit de regles de decisió creades a partir d'un coneixement previ, pres com a cert. En el nostre sistema de determinació, si ens limitem a aquesta funció farem ús de l'anàlisi discriminant o tècniques relacionades. Hom hauria de poder modificar les regles discriminants per tal de millorar les seves respostes i adequar-les més al criteri de l'expert. En el segon cas, en canvi, el sistema permet posar en dubte el mateix coneixement científic en el que es basa l'expert. Per tant, resulta útil tant per a l'usuari inexpert com per a l'expert. Aquest segon model es fa especialment necessari en casos on el coneixement expert no està del tot acceptat o definit. L'usuari inexpert tan sols utilitza el sistema de determinació per a avaluar inventaris, mentre que l'usuari expert emprava alhora el sistema com a una ajuda per a avaluar el seu coneixement de la matèria. En relació amb la capacitat d'aprenentatge del sistema cal notar que la modificació del cos de coneixement del sistema de determinació es podria automatitzar a partir de la inclusió, en aquest cos de coneixement, dels nous inventaris avaluats que complissin uns determinants requisits. No obstant, aquesta opció comporta el risc de corrompre les unitats de vegetació quan els inventaris avaluats tenen característiques aberrants o intermèdies. Per tant la modificació del cos de coneixement ha de romandre reservada al expert a càrrec del sistema. En aquest cas la modificació del cos de coneixement no és automàtica sinó semiautomàtica.

En els propers apartats descrivim el sistema de determinació *Araucaria*, una aplicació client-servidor per a la determinació automàtica d'inventaris de vegetació. Per a començar, expliquem el tractament dels inventaris previ a la seva introducció al sistema. Tot seguit descrivim les propietats de la definició d'unitat de classificació que hem adoptat i justifiquem la agrupació d'unitats de classificació en àrees de classificació. Finalment, descrivim com es creen i gestionen les unitats de classificació dins del sistema de determinació.

### 4.1.2.2 Importació i pretractament de les dades

Ja vam comentar al capítol 1.2 la necessitat d'homogeneïtzació de la nomenclatura quan es tracten inventaris d'origens diversos. La homogeneïtzació de la nomenclatura es basa en la informació continguda en un tesaure de tàxons que sigui un reflex de la nomenclatura taxonòmica acceptada per un equip de botànics. Per tant, la selecció del tesaure taxonòmic és el primer pas en la configuració del sistema.

La importació d'inventaris de vegetació al sistema de determinació incorpora un protocol de pretractament de les entrades dels tàxons, que resumim a la taula 4.1.1. A més de la homogeneïtzació de la nomenclatura durant la importació d'inventaris el sistema realitza altres tasques com, per exemple, l'eliminació d'entrades amb determinació dubtosa o la suma de entrades corresponents a un tàxon que apareix a diferents estrats. Totes aquestes operacions estan activades per defecte a *Araucaria*, però poden ser desactivades. El mateix protocol de pretractament és el que s'aplica a les noves taules d'inventaris que el sistema rep per a donar una resposta de determinació a l'usuari no expert.

L'inventari finalment emmagatzemat es compon simplement d'una llista de noms de tàxons vàlids, associats a un camp d'abundància, ja sigui una classe de cobertura-abundància o un valor d'abundància numèric. La resta d'informació associada a l'inventari és descartada per ésser innecessària per al motor de funcionament del sistema. Els inventaris es troben agrupats en el que anomenarem **seccions de dades**. Cada secció de dades permet importar taules d'inventaris en format Quercus (vegeu capítol 4.2) i realitzar el pretractament de les dades esmentat més amunt. Aquesta segmentació en facilita la gestió, perquè cada sintàxon de rang alt (classe, ordre, ...) pot ésser responsabilitat de diferents experts.

<b>Acció</b>	<b>Valor per defecte</b>
Eliminar entrades de determinació dubtosa (cf.)	Activat
Eliminar entrades de tàxons no reconeguts al tesaure	Activat
Eliminar entrades tàxons supra-específics (p.ex. <i>Festuca</i> sp.)	Activat
Sumar entrades observades en estrats diferents (p.ex. arbustiu i arbori)	Activat
Sumar entrades amb certeses de determinació diferents (p.ex. sensu lato o grup)	Activat
Traducció de noms sinònims a noms vàlids	Activat

**Taula 4.1.1:** Protocol de tractament dels inventaris importats al sistema de determinació *Araucaria* o que han de ser sotmesos a una determinació. Les diferents accions poden ésser activades/desactivades dins del programa.

#### 4.1.2.4 Unitats de *clustering* independents (UCI)

La definició del concepte de clúster és un punt crític d'un sistema de determinació que aspiri a resultar d'utilitat. En primer lloc, la definició de clúster ha d'ésser un reflex del concepte de unitat de vegetació d'ecòlegs i fitosociòlegs (veure capítol 3.1). A la vegada, la definició ha d'ésser prou senzilla per tal que un usuari no expert pugui interpretar amb facilitat els resultats de la seva consulta al sistema. A partir de raonaments exposats en capítols anteriors, les característiques que han de presentar les unitats de vegetació són:

1. El clúster és representat pel centroide geomètric del conjunt d'inventaris que en formen part.
2. Les relacions espacials entre inventaris es basen exclusivament en la composició florística (taxonòmica) de la comunitat. El centroide representa doncs, la composició florística mitjana del clúster.
3. L'espai de relacions entre inventaris ha d'excloure les dobles absències. Entre les mesures comparades al capítol 3.2 utilitzarem la distància de Hellinger (Rao 1995) o la distància de Bray-Curtis (1957) que són dues de les que han donat millors resultats.
4. El conjunt d'inventaris que defineix el clúster és difús (*fuzzy*): Cada inventari expressa amb més o menys força el concepte associat a la unitat de vegetació segons el seu grau de pertinença (*membership*). L'exponent de *fuzziness* ( $m$ ) és un paràmetre del sistema que cal adequar a la "borrositat" de les dades.
5. Les unitats de *clustering* són independents, no formen part d'una partició. Aquest fet implica que un inventari pot restar sense classificar o expressar característiques de diferents unitats de vegetació a la vegada. Ja vàrem discutir al capítol 3.3 alguns problemes associats a les particions. A la pràctica presenten l'inconvenient afegit de necessitar iniciar de nou el procés de *clustering* de tots els grups quan només es vol alterar el coneixement d'una sola unitat. La consideració de clústers difusos independents ens porta al model de clúster possibilístic i a la seva implementació en l'algorisme *Possibilistic C-Means* (PCM, Krishnapuram & Keller 1993). A partir de l'algorisme *meta-PCM* descrit al capítol 3.3 podem estimar la distància de referència adequada per a cada clúster.
6. Hom podria optar per incloure una ponderació dels tàxons en la definició de clúster. En vista dels resultats del capítol 3.4 hem descartat aquesta opció, que a la vegada suposava un cost addicional en temps de computació important.

Resumint, les mesures de proximitat per a relacionar inventaris són les distàncies de Bray-Curtis o Hellinger, la definició de clúster adoptada és la proporcionada per l'algorisme *Possibilistic C-means*, i el mètode de generació de clústers el proporcionat per l'algorisme *meta-PCM*.

Les unitats de classificació poden ser actives o inactives. Les primeres són clústers de *PCM* optimitzats mitjançant l'algorisme *meta-PCM*. Les unitats inactives són grups d'inventaris que l'expert estableix com a clúster immutable, per a ésser emprat només en la determinació d'inventaris externs. Evidentment, aquestes unitats no estan suportades estadísticament per l'algorisme *meta-PCM*. Per aquesta raó no es recomana abusar de la definició d'unitats inactives ja que no permeten avaluar el coneixement de l'expert i degraden la capacitat de determinació del sistema.

Entre els atributs d'una UCI més importants, hi ha:

- Nom i descripció de la unitat.
- Sintàxon de referència: Sintàxon representat per la unitat, si és que aquest està definit. L'associació entre un sintàxon i una UCI es fa a través d'un tesaure de sintàxons. Aquesta associació posa la unitat de vegetació en un context sintaxonòmic, permetent conèixer la sintaxonomia de nivells superiors a la unitat. Si el sintàxon té un inventari tipus, aquest ha de formar part dels inventaris membres de la UCI.
- L'estat de la unitat: Escoltant, en procés de relaxació, determinant un inventari, o inactiu.
- La distància de referència: paràmetre de l'algorisme *PCM* estimat a través de l'algorisme *meta-PCM*.
- Cardinalitat i variabilitat geomètrica: Indicadors de la mida i variabilitat del clúster, respectivament.
- Primera derivada: Valor mínim local, trobat per la darrera execució de *meta-PCM*.
- Segona derivada: Informa de si el clúster està optimitzat segons el criteri de l'algorisme *meta-PCM*. Per a clústers optimitzats el seu valor hauria de ser proper a zero.

Donada una configuració del clúster les accions de *clustering* que es poden donar en una unitat activa són tres:

- a. Relaxament del clúster:** Executa *meta-PCM* (vegeu el quadre 3.3.1) per a trobar la configuració del clúster i la distància de referència que corresponen a un mínim de la primera derivada de la variabilitat geomètrica.
- b. Creixement del clúster:** Executa *meta-PCM* incrementant inicialment la distància de referència fins a sobrepassar el màxim de la primera derivada de la variabilitat geomètrica. Seguidament relaxa el clúster per a assolir el mínim següent. El procés sencer permet augmentar la mida del clúster fins a la propera mida a la que es detecta un node dens.
- c. Reducció del clúster:** Executa *meta-PCM* disminuint inicialment la distància de referència fins a sobrepassar el màxim de la primera derivada de la variabilitat geomètrica. Seguidament relaxa el clúster per a assolir el mínim següent. El procés sencer permet disminuir la mida del clúster per a trobar a un subgrup dens.

#### 4.1.2.5 Àrees de classificació

De manera anàloga a l'agrupació d'inventaris en seccions, les UCIs que es defineixen en el sistema *Araucaria* es troben agrupades en el que anomenarem **àrees de classificació**. Una àrea de classificació constitueix el cos de coneixement corresponent als sintàxons de base d'un tipus de vegetació (normalment un sintàxon de nivell sintaxonòmic alt: aliança, ordre o classe). La divisió del conjunt d'unitats de classificació en àrees de classificació permet que diferents experts es puguin fer responsables de diferents àrees, sense que l'acció sobre un tipus de vegetació d'un d'ells modifiqui les decisions d'anàlisi preses pels altres en les seves respectives àrees de coneixement.

A cada àrea de classificació els inventaris sobre els que s'estableix la classificació són diferents. Són els **inventaris actius** de l'àrea de classificació. Sobre l'elecció dels inventaris a activar en una àrea de classificació cal remarcar que és un pas decisiu pel bon funcionament del sistema. En una base de dades d'inventaris és freqüent la aparició d'inventaris amb característiques intermèdies. La inclusió indiscriminada de tots els inventaris de la base de dades pot conduir a que els inventaris intermedis dificultin o impedeixin la detecció de les unitats de vegetació presents. En el millor dels casos, es inevitable que en modifiquin el centroide del clúster i, per tant, la definició de la unitat. És important, doncs, que el conjunt d'inventaris actius d'una àrea de classificació inclogui els inventaris més típics de les unitats de vegetació a representar, i romanguin exclosos aquells inventaris considerats intermedis.

Quan el sistema posa en marxa el motor d'una àrea de classificació, es llegeixen els inventaris associats i es transformen segons les especificacions de la mateixa (sinonímia d'anàlisi, transformacions...). A partir dels inventaris transformats, a l'àrea de classificació es construeix una matriu de dissimilaritats entre inventaris. Aquest espai multidimensional de proximitats és el que el sistema finalment utilitza per a establir els clústers d'inventaris de vegetació. Entre les diferents àrees poden variar aspectes relatius a la transformació dels inventaris i la construcció de l'espai de proximitats (vegeu la taula 4.1.2).

<b>Característica</b>	<b>Valor per defecte</b>
Aplicació d'una sinonímia d'anàlisi: tractar tàxons diferents com a entitats d'anàlisi equivalents (és a dir, considerar que especifiquen la mateixa variable).	Variable.
Traducció de l'escala ordinal de Braun-Blanquet a una escala enterament numèrica	Transform. combinada (van der Maarel 1979)
Transformació numèrica escalar (p.ex. $y = \sqrt{x}$ , o $y = \log(x+1)$ )	Desactivat
Mesura de proximitat (dissimilaritat) entre inventaris de vegetació. Actualment l'expert pot escollir entre les distàncies de Bray-Curtis (1957) i Hellinger (Rao 1995).	Distància de Bray-Curtis

**Taula 4.1.2:** Atributs que conformen la configuració de les àrees de classificació. Per a les diferents opcions de cada característica vegeu el capítol 3.2.



#### 4.1.2.6 Creació i gestió d'UCIs en una àrea de classificació

El sistema de determinació proporciona una senzilla eina per a inicialitzar UCIs, que permet explorar l'espai multidimensional de proximitats. Aquesta eina segueix el següent protocol d'actuació:

1. A partir dels inventaris actius de la àrea encara sense classificar es proposa a l'expert una tríada d'inventaris per a actuar com a llavors de *clustering*.
2. Es calculen les distàncies entre tots els inventaris actius i el centroide de la tríada.
3. La distància de referència inicial es calcula com aquella que proporciona, segons l'equació de tipicalitats de *PCM*, una cardinalitat del clúster de 3 (tríada).
4. L'usuari pot llavors iniciar el creixement de la unitat diverses vegades, trobant configuracions de clústers cada cop més grans, fins a la mida desitjada. Qualsevol d'aquests estadis compleix la condició de correspondre a un mínim de la primera derivada de la variabilitat geomètrica.
5. L'usuari finalment accepta o rebutja el clúster i la eina de classificació proporciona una nova tríada d'inventaris per a una nova unitat.

En el procés de creació d'una àrea de *clustering* cal alternar la activació d'inventaris corresponents a una unitat de vegetació i la creació d'una UCIs que la representi. És important remarcar que la activació de nous inventaris provoca el relaxament de les unitats de clúster actives preexistents. Es pot donar el cas que, en relaxar-se, dues UCIs arribin a imbricar-se tant com per a esdevenir llavors unitats iguals. Aquest fet és més probable quan els inventaris que s'activen a l'àrea de classificació corresponen a inventaris intermedis. És evident que l'encavalcament entre UCIs es un fet que l'expert ha d'evitar. Per aquest motiu, *Araucaria* proporciona eines per a detectar-ho, per exemple, mitjançant la inspecció d'una matriu d'encavalcaments entre conjunts borrosos o amb l'anàlisi de relacions de distància entre centroides.

És possible definir en una àrea unitats de classificació inactives. En la creació d'una unitat de clúster inactiva el sistema estima una distància de referència per la unitat de la mateixa manera que el pas (1) de l'exploració d'unitats actives. Aquesta distància no té efecte sobre el còmput de les pertinences relatives dels inventaris a determinar però sí sobre el càlcul de tipicalitats.

#### 4.1.2.7 Interfície de l'usuari expert

A la figura 4.1.1 mostrem una vista de la interfície de l'usuari expert del sistema *Araucaria*. Aquesta interfície es compon de diverses finestres:

1. La finestra del projecte (a dalt a l'esquerra) conté la llista d'àrees de classificació, la llista de seccions de la base de dades, i els tesausres taxonòmic i sintaxonòmic que el sistema utilitza.
2. Cada secció de dades s'edita en una finestra independent, des d'on es poden importar inventaris en format *xml* de *Quercus* (capítol 4.2).
3. De manera anàloga, cada àrea de classificació es pot editar en una finestra independent (a la dreta). Aquesta finestra conté la llista d'UCIs definides l'àrea i altres informacions, com la llista d'inventaris activats, la matriu d'encavalcament entre clústers o la llista de sinònims d'anàlisi.
4. A partir de la llista d'unitats de l'àrea és possible obrir un diàleg per a treballar amb la definició d'una UCI determinada (a baix a l'esquerra).

The screenshot displays the 'Araucaria: Sistema de classificació automàtica de la vegetació - NewAraucaria.zip' application window. The interface is divided into several panes:

- Centre del projecte:** Shows 'Seccions d'inventaris: 2' and 'Àrees de clustering: 4'. It includes tabs for 'Tesaure de taxonomia', 'Tesaure de sintaxonomia', 'Àrees de clustering', and 'Seccions de la base de dades'. A table lists areas like 'Xerobromion erecti' with 12 clusters.
- Cluster: AB:** A dialog box for editing cluster 'AB', showing 'Estat: Escoltant', 'Fuzziness(m) = 1,05', 'R.D. = 0,478', and 'Card.=17,755'. It has tabs for 'Inventaris', 'Tàxons', 'Descripció', and 'Output'.
- Area: Xerobromion erecti:** A detailed view of the 'Xerobromion erecti' area, showing 'Inventaris actius: 145' and 'Nº clústers: 12'. It includes a table of clusters with columns for 'Clúster', 'Estat', 'Dist. Ref.', 'Cardina...', 'Geom...', 'dVG', and 'ddVG'.
- Server Window:** A small window in the top right corner showing 'Waiting...'.

Figura 4.1.1: Vista de la interfície del sistema de determinació *Araucaria*.

#### 4.1.2.8 Avaluació d'inventaris de vegetació a través d'una aplicació client

El sistema de determinació *Araucaria* proporciona un servei remot de determinació automàtica d'inventaris de vegetació. A través d'aquest accés remot, un usuari no expert pot enviar per la xarxa una taula d'inventaris de vegetació per a demanar al sistema que l'avaluï i recollir-ne els resultats. La interacció usuari-sistema es realitza mitjançant una aplicació client, de la qual en mostrem una vista a la figura 4.1.2. A continuació descrivim els passos que s'esdevenen en el procés de l'avaluació d'una taula d'inventaris des de l'aplicació client d'*Araucaria*:

1. Càrrega del fitxer de la taula d'inventaris en format .xml: El fitxer de dades ha de complir el format *xml* de QUERCUS (veure capítol 4.2).
2. Pretractament de la taula d'inventaris: El pretractament que reben els inventaris a determinar segueix el mateix protocol que seguit en la incorporació d'inventaris a la base de dades del sistema (vegeu apartat 4.1.1.3). L'usuari envia la taula d'inventaris i rep una nova taula tractada, juntament amb una sortida de text on s'enumeren accions realitzades en el procés de pretractament. L'aplicació client envia de nou la taula d'inventaris tractada al servidor, aquest cop per a ésser avaluada.
3. Determinació dels inventaris: En l'avaluació de la taula tractada, el sistema de determinació efectua les següents operacions:
  - 3.1. Per a cada inventari el sistema calcula la distància al centroides de cada UCI definida al sistema de determinació. Aquestes distàncies constitueixen la base geomètrica sobre la qual el sistema de determinació construeix la seva resposta. Les UCIs s'ordenen per distàncies creixents i es guarden només aquelles unitats més properes.
  - 3.2. Sumant aquelles unitats més properes als inventaris de la taula s'obté la llista d'UCIs que finalment serveixen per realitzar la determinació. És important reduir el nombre final d'UCIs per simplificar la resposta que obté l'usuari de l'aplicació client.
  - 3.3. Càlcul de la tipicalitat (pertinença absoluta) de cada inventari a cada UCI, a partir de la comparació entre la distància de l'inventari al centroides del clúster i la distància de referència de la unitat. Un inventari amb una tipicalitat alta en una unitat és un bon candidat per augmentar la base de coneixement del sistema si l'expert encarregat de l'àrea de classificació així ho creu convenient.
  - 3.4. Càlcul de les pertinences relatives a partir de les distàncies a tots els centroides i tenint en compte un exponent de *fuzziness*. Les pertinences relatives d'un inventari a cada un dels clústers de l'àrea sumen 1. Per tant, és possible interpretar aquests valors com probabilitats, tot que el concepte de pertinença difusa es diferent. El clúster per al qual l'inventari presenta una pertinença relativa més alta (és a dir, el que presenta el centroides més proper) és la unitat de vegetació més probable des del punt de vista del sistema.

4. Resultat de la determinació: El servidor del sistema de determinació retorna a la aplicació client informació sobre l'avaluació de la taula d'inventaris. Aquesta informació consta de: (1) les matrius de distàncies, tipicalitats i pertinences relatives; (2) una descripció del sintàxon associats a les UCIs; (3) el llistat de la constància i fidelitat dels tàxons presents a cada clúster; i (4) la una sortida de text enunciant les accions dutes a terme durant la determinació de la taula. Totes aquestes dades, així com la informació obtinguda en el procés de pretractament pot ésser guardada en fitxers si l'usuari de l'aplicació client així ho desitja.

A l'adreça d'Internet <http://biodiver.bio.ub.es/vegana/araucaria/> hom hi pot trobar l'enllaç per al programa client, així com unes breus instruccions d'ús i una enumeració dels sintàxons als quals el sistema *Araucaria* dona resposta.

The screenshot shows the 'Araucaria: Eina de determinació d'inventaris de vegetació' application window. The main area displays a table with columns for 'Pertinences relatives', 'Distàncies', and 'Tipicalitats'. The 'Pertinences relatives' column is highlighted in blue. Below the table, there is a section for 'Clúster: AB Syntaxon: Adonido-Brometum erecti' which includes a 'Syntaxonomy' tree and a list of taxa with their 'Presen...' and 'Phi' values.

	AB	AD1	AD2	IBL/IBT	KAT/KAA	KG	LB	TAV	TBF	TBH
p-p00091	0,77	0,03	0,01	0,02	0,01	0,02	0,04	0,02	0,00	0,04
p-p00092	0,68	0,01	0,01	0,02	0,01	0,02	0,14	0,01	0,00	0,05
p-p00093	0,92	0,00	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,01
p-p00094	0,90	0,00	0,00	0,01	0,01	0,04	0,01	0,00	0,00	0,01
p-p00095	0,67	0,07	0,03	0,07	0,01	0,03	0,03	0,01	0,01	0,05
p-p00096	0,94	0,01	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,01
p-p00097	0,97	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01
p-p00098	0,98	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p-p00099	0,98	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
p-p00100	0,80	0,02	0,01	0,05	0,05	0,01	0,01	0,00	0,00	0,04
p-p00101	0,98	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
p-p00102	0,94	0,00	0,00	0,01	0,02	0,00	0,00	0,00	0,00	0,01
p-p00103	0,95	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,01
p-p00104	0,95	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,02

Clúster: AB Syntaxon: Adonido-Brometum erecti

Syntaxonomy:

- .Class Festuco-Brometea Br.-Bl. et R. Tx. 1943
- .Order Brometalia erecti Br.-Bl. 1936
- .Alliance Xerobromion erecti (Br.-Bl. et Moor) R. Tx. et Oberd. 1958
- .Sub-alliance Eu-Xerobromenion Oberd. 1957
- .Association Adonido-Brometum erecti Font 1983

Taxon	Presen...	Phi
Adonis vernalis L.	12,9212	0,7983
Onobrychis supina (Vill.) DC. in Lam. ...	17,7134	0,5131
Seseli montanum L.	10,0027	0,0345
Phleum phleoides (L.) Karsten	7,0482	0,003
Odontides verna (Bellardi) Dumort.	2,9688	-0,0908
Teucrium pyrenaicum L.	4,7523	-0,0451
Helianthemum nummularium (L.) Mill.	14,765	0,1097
Euphorbia cyparissias L.	13,8614	0,2497
Scabiosa columbaria L.	12,9465	0,1134
Bromus erectus Huds.	8,9576	0,0742
Plantago media L.	15,9304	0,3921

**Figura 4.1.2:** Vista de l'aplicació client d'*Araucaria*, un cop rebuts els resultats de la determinació d'una taula d'inventaris. A la part superior apareix la matriu de pertinences relatives, on les UCIs corresponen a les columnes. En seleccionar una columna a la part inferior apareix la descripció del sintàxon associat i la llista de constàncies i fidelitats dels tàxons de la UCI corresponent.

### 4.1.3 Construcció i avaluació de les àrees de *Brometalia erecti* i *Quercetea ilicis*

#### 4.1.3.1 Introducció de *Brometalia erecti* i *Quercetea ilicis* al cos de coneixement del sistema

L'objectiu d'aquest apartat és la creació d'un cos de coneixement inicial per al sistema de determinació *Araucaria*. Per a assolir-lo, en primer lloc s'han introduït al sistema dues taules d'inventaris, una amb els 531 inventaris de l'ordre *Brometalia erecti* i l'altra amb els 753 inventaris de la classe *Quercetea ilicis*. Cada una d'aquestes taules d'inventaris ha constituït una secció de dades del sistema. A continuació, s'han creat les quatre àrees de classificació següents:

Àrea 1: Al. *Xerobromion erecti*

Àrea 2: Al. *Mesobromion erecti*

Àrea 3: Cl. *Quercetea ilicis* sense O. *Quercenion ilicis*

Àrea 4: O. *Quercenion ilicis*

La configuració de les àrees de classificació ha estat:

- a) Els sinònims d'anàlisi ja havien estat tractats a cada una de les taules d'inventaris (vegeu el capítol 1.2).
- b) Transformació combinada (van der Maarel 1979) de l'escala ordinal de Braun-Blanquet.
- c) Transformació escalar amb exponent  $w = 1$  (àrees 1 i 2) i  $w = 1.5$  (àrees 3 i 4). Hom pot trobar la raó d'aquesta elecció al capítol 3.2.
- d) La mesura de proximitat escollida ha estat la distància de Hellinger (Rao 1995).
- e) Coeficient de *fuzziness* ( $m$ ) = 1.05 (vegeu el capítol 3.3).

La construcció de les unitats de *clustering* independents (UCIs) de cada àrea de classificació s'ha fet seguint el protocol descrit a l'apartat 4.1.2.6. A les següents taules 4.1.3.A-B i 4.1.3.B-C mostrem els atributs de cada unitat de classificació creada. La darrera columna indica el sintàxon al que la unitat representa.

La construcció de les UCIs ha intentat conservar, sempre que això ha estat possible, la correspondència 1 a 1 amb els sintàxons de base, és a dir, una UCI per a cada sintàxon de base. Per a alguns sintàxons, però, la unitat construïda engloba diferents subassociacions de la mateixa associació. És el cas de les associacions *Irido-Brometum erecti* o *Koelerio-Avenuletum ibericae*. Especialment difícil és la distinció en unitats diferents de les subassociacions de garrigars (*Quercetum cocciferae*) i de l'associació *Rhamno-Quercetum cocciferae*, que han estat finalment englobades en una sola unitat: QC/RQ. Per contra, en els sintàxons de base *Euphrasio-Plantaginetum* subass *typicum*, *Achilleo-Dichanthietum ischaemi* i *Oleo-Lentiscetum*, s'han creat dues UCIs per a cada sintàxon de base, per considerar-se massa variables. Aquest fet fa palesa la gran variabilitat territorial de la unitat de vegetació, cosa per la qual el sistema de

determinació treballa millor si es defineixen dues UCIs enlloc d'una sola, malgrat el sintàxon associat a les dues sigui el mateix.

La majoria d'UCIs actives definides tenen una derivada segona de la variabilitat geomètrica propera a zero, indicant una optimització segons el criteri de *meta-PCM*. La cardinalitat és relativament baixa en força casos, sobretot en aquelles unitats molt properes a altres unitats.

En aquells sintàxons en que no era possible configurar una UCI estable, s'ha optat per generar una unitat inactiva (que a les taules 4.1.3.B i 4.1.3.D s'indica mitjançant un asterisc '\*')

#### A. Àrea 1: *Al. Xerobromion erecti*

	<i>RD</i>	<i>Card</i>	<i>VG</i>	<i>dVG</i>	<i>ddVG</i>	<i>Sintàxon de referència</i>
<b>AB</b>	0.7771	17.56	0.4701	0.3457	-0.0001	<i>Adonido-Brometum erecti</i>
<b>AD1</b>	0.6967	8.95	0.3358	0.0698	-0.0005	<i>Achilleo-Dichanthietum ischaemi</i>
<b>AD2</b>	0.7610	5.99	0.3152	0.0505	-0.0005	<i>Achilleo-Dichanthietum ischaemi</i>
<b>IB</b>	0.7456	5.96	0.3749	0.1122	-0.0011	<i>Irido-Brometum erecti</i>
<b>KA</b>	0.7680	12.61	0.4359	0.5197	0.0016	<i>Koelerio-Avenuletum ibericae</i>
<b>KG</b>	0.7106	7.92	0.3341	0.2007	-0.0018	<i>Koelerio-Globularietum punctatae</i>
<b>LB</b>	0.6604	5.93	0.3329	0.0505	-0.0024	<i>Lino viscosi-Brometum erecti</i>
<b>TAV</b>	0.7864	11.97	0.3854	0.0368	0.0002	<i>Teucrio-Avenuletum mirandanae</i>
<b>TBF</b>	0.8852	12.97	0.4771	0.0481	-0.0007	<i>Teucrio-Brometum erecti subass festucetosum fallacis</i>
<b>TBH</b>	0.6989	5.95	0.3721	0.0373	-0.0004	<i>Teucrio-Brometum erecti subass helianthemetosum</i>
<b>TBT</b>	0.7211	5.96	0.3763	0.0451	0.0001	<i>Teucrio-Brometum erecti subass typica</i>
<b>TF</b>	0.8207	5.99	0.3801	0.0416	0.0005	<i>Teucrio-Festucetum spadiceae</i>

#### B. Àrea 2: *Al. Mesobromion erecti*

	<i>RD</i>	<i>Card</i>	<i>VG</i>	<i>dVG</i>	<i>ddVG</i>	<i>Sintàxon de referència</i>
<b>AF</b>	0.6608	5.98	0.3140	0.0174	-0.0003	<i>Alchemillo-Festucetum nigrescentis</i>
<b>AP</b>	0.7086	8.97	0.3311	0.0411	-0.0005	<i>Astragalo danici-Poetum alpinae</i>
<b>CAT</b>	0.7004	12.83	0.3363	0.2274	0.0005	<i>Chamaespartio-Agrostidietum subass typica</i>
<b>CB*</b>	0.4585	8.00	0.1687	0.2008	-5.0985	<i>Carlino-Brachypodietum pinnati</i>
<b>CG</b>	0.6947	7.98	0.2884	0.0505	-0.0002	<i>Centaureo-Genistetum tinctoriae</i>
<b>EPB*</b>	0.3940	5.00	0.1215	0.1141	-8.9487	<i>Euphrasio-Plantaginetum mediae subass brachypodietosum</i>
<b>EPC</b>	0.7603	7.96	0.3855	0.0716	0.0007	<i>Euphrasio-Plantaginetum mediae subass centauretosum</i>
<b>EPG</b>	0.6150	3.99	0.2356	0.0376	0.0007	<i>Euphrasio-Plantaginetum mediae subass gentianelletosum</i>
<b>EPT1</b>	0.7972	4.00	0.3442	0.0126	0.0003	<i>Euphrasio-Plantaginetum mediae subass typicum</i>
<b>EPT2</b>	0.6399	4.00	0.2528	0.0042	0.0007	<i>Euphrasio-Plantaginetum mediae subass typicum</i>
<b>GP</b>	0.7363	10.99	0.3287	0.0141	-0.0007	<i>Gentiano acaulis-Potentilletum montanae</i>
<b>GT</b>	0.7208	11.93	0.3435	0.0968	-0.0004	<i>Gentiano-Trifolietum montani</i>
<b>KT</b>	0.9064	7.99	0.4512	0.0350	-0.0012	<i>Koelerio-Trifolietum molinerii</i>
<b>PF</b>	0.7244	5.90	0.3962	0.1581	0.0032	<i>Phyteumo orbicularis-Festucetum nigrescentis</i>
<b>PS</b>	0.7731	4.94	0.4488	0.0819	-0.0012	<i>Plantagini-Seslerietum</i>
<b>TAS</b>	0.6854	6.92	0.3332	0.1969	0.0016	<i>Teucrio pyrenaici-Astragaletum catalaunici</i>

**Taules 4.1.3.A-B:** Atributs de les unitats de *clustering* independents (UCIs) creades per a les àrees 1 i 2. **RD:** distància de referència **Card.:** Cardinalitat. **VG:** Variabilitat geomètrica. **dVG:** Primera derivada de la variabilitat geomètrica. **ddVG:** Segona derivada de la variabilitat geomètrica. (\* = unitats de *clustering* inactives).

**C. Àrea 3: CI. *Quercetea ilicis* sense *O.Quercenion ilicis***

	<i>RD</i>	<i>Card</i>	<i>VG</i>	<i>dVG</i>	<i>ddVG</i>	<i>Sintàxon de referència</i>
<b>BJ</b>	0.7808	17.98	0.3268	0.0417	-0.0009	<i>Buxo sempervirentis-Juniperetum phoeniceae</i>
<b>CM</b>	0.5850	4.00	0.1871	0.0063	0.0005	<i>Calicotomo-Myrtetum</i>
<b>CO</b>	0.7892	14.98	0.4068	0.0104	-0.0009	<i>Clematido-Osyrietum albae</i>
<b>MJ</b>	0.6922	10.00	0.2671	0.0048	0.0008	<i>Myrto-Juniperetum oxycedri</i>
<b>OL1</b>	0.7641	5.00	0.2731	0.0003	0.0010	<i>Oleo-Lentiscetum provinciale</i>
<b>OL2</b>	0.7570	4.99	0.3084	0.0342	-0.0006	<i>Oleo-Lentiscetum provinciale</i>
<b>QC/RQ</b>	0.7188	23.83	0.3308	0.1523	0.0000	<i>Quercetum cocciferae/Rhamno-Quercetum cocciferae</i>
<b>QL</b>	0.6040	19.89	0.2815	0.3407	-8.4561	<i>Querco-Lentiscetum</i>
<b>QRB</b>	0.7109	16.92	0.3618	0.0464	0.0010	<i>Quercetum rotundifoliae subass buxetosum</i>
<b>QRR</b>	0.6358	4.99	0.2846	0.0100	0.0002	<i>Quercetum rotundifoliae subass rhamnitosum infect.</i>
<b>QRU</b>	0.7100	12.78	0.3676	0.2698	0.0038	<i>Quercetum rotundifoliae subass ulicetosum</i>
<b>RJ</b>	0.7084	6.99	0.3273	0.0193	-0.0009	<i>Rhamno lycioidis-Juniperetum phoeniceae</i>

**D. Àrea 4: *O. Quercenion ilicis***

	<i>RD</i>	<i>Card</i>	<i>VG</i>	<i>dVG</i>	<i>ddVG</i>	<i>Sintàxon de referència</i>
<b>AQB</b>	0.6627	10.99	0.2517	0.0212	-0.0009	<i>Asplenio-Quercetum ilicis subass buxetosum</i>
<b>AQHB</b>	0.5879	5.00	0.2106	0.0024	0.0009	<i>Asplenio-Quercetum ilicis subass hilocomio-buxet.</i>
<b>AQS</b>	0.6734	4.99	0.3041	0.0232	0.0017	<i>Asplenio-Quercetum ilicis subass sarothamnetosum</i>
<b>AQT</b>	0.6161	4.00	0.2338	0.0002	0.0010	<i>Asplenio-Quercetum ilicis subass typicum</i>
<b>AQTL</b>	0.6453	4.94	0.3267	0.0256	0.0008	<i>Asplenio-Quercetum ilicis subass torminalo-ligustretosum</i>
<b>CYB</b>	0.5418	5.99	0.1939	0.0127	-0.0009	<i>Cytiso-Bupleuretum fruticosi</i>
<b>VQAC</b>	0.7048	4.00	0.2943	0.0033	0.0008	<i>Viburno-Quercetum ilicis subass aceretosum monspesulani</i>
<b>VQAR*</b>	0.7200	5.00	0.3472	1.7518	19.0872	<i>Viburno-Quercetum ilicis subass arbutetosum</i>
<b>VQAS</b>	0.5850	4.00	0.1517	0.0000	0.0004	<i>Viburno-Quercetum ilicis subass asplenietosum</i>
<b>VQC</b>	0.6306	6.99	0.2517	0.0247	-0.0006	<i>Viburno-Quercetum ilicis subass cerroidetosum</i>
<b>VQE</b>	0.7156	8.99	0.3103	0.0085	0.0006	<i>Viburno-Quercetum ilicis subass ericetosum</i>
<b>VQP*</b>	0.7550	9.00	0.5012	1.2276	24.5203	<i>Viburno-Quercetum ilicis subass pistacietosum</i>
<b>VQQP</b>	0.7173	7.00	0.2923	0.0029	0.0009	<i>Viburno-Quercetum ilicis subass quercetosum pubescent.</i>
<b>VQS</b>	0.6964	18.25	0.3683	0.3793	0.0000	<i>Viburno-Quercetum ilicis subass suberetosum</i>
<b>VQV</b>	0.7122	3.99	0.3526	0.0211	0.0001	<i>Viburno-Quercetum ilicis subass viburnetosum</i>

**Taules 4.1.3.C-D:** Atributs de les unitats de *clustering* independents (UCIs) creades per a les àrees 3 i 4. **RD:** distància de referència **Card.:** Cardinalitat. **VG:** Variabilitat geomètrica. **dVG:** Primera derivada de la variabilitat geomètrica. **ddVG:** Segona derivada de la variabilitat geomètrica. (\* = unitats de *clustering* inactives).

#### 4.1.3.2 Avaluació de la capacitat de resposta del sistema de determinació

En aquest apartat ens proposem testar la capacitat de resposta del sistema de determinació. Concretament hem avaluat la resposta del sistema separatament per a cada una de les àrees de classificació introduïdes a l'apartat anterior. Hem realitzat l'avaluació de cada àrea comparant la UCI per la que cada inventari presentava el grau de pertinença relatiu més alt amb el sintàxon de base al que l'inventari pertanyia. En el cas d'inventaris prèviament emprats en la construcció de l'àrea de classificació la determinació s'ha fet extraient primer l'inventari i relaxant aquelles UCIs que ho requerien. Per tant, es pot considerar una validació creuada per *leave-one-out*. En el cas dels inventaris inactius, és a dir, aquells inventaris descartats en la construcció de l'àrea de classificació, la determinació s'ha realitzat com si fossin inventaris nous.

Cal admetre que en aquest estudi d'avaluació hem suposat l'àrea de classificació a la que es determinava l'inventari com a coneguda. Per tant, no hem avaluat la capacitat del sistema per a determinar correctament l'àrea de classificació a la que cada inventari pertany. No obstant, no creiem que aquest fet hagi afectat en gran mesura els resultats.

Per tal de presentar els resultats de l'avaluació de la capacitat de predicció del sistema de determinació mostrem quatre taules de confusió (taules 4.1.4.A-D) corresponents a la comparació, en cada àrea de classificació, entre el sintàxon de base original (en columnes) i la UCI determinada pel sistema (en files). Hem ressaltat aquelles combinacions que hem considerat encerts. Lògicament, la majoria dels encerts corresponen a caselles de la diagonal de les matrius. En els casos en que s'han construït UCIs que engloben més d'un sintàxon de base o en que dues UCIs diferents assenyalen el mateix sintàxon de base les caselles considerades encerts tenen elements fora de la diagonal. Per simplicitat, les matrius de confusió mostrades engloben tant la determinació dels inventaris actius, a través de la tècnica de validació creuada *leave-one-out*, com la determinació dels inventaris inactius.

Quan es produeix un error de determinació, el clúster proposat pel sistema és normalment una unitat propera a la considerada correcte segons el criteri sintaxonòmic tradicional. Per exemple, els inventaris de l'associació *Lino-Brometum erecti* que el sistema no determina dins de la unitat LB, van a parar sobretot a UCIs de *Teucrio-Brometum erecti* associació inclosa dins de la mateixa subaliança *Eu-Xerobromenion*. De manera semblant, els inventaris de *Rhamno-Quercetum cocciferae* (RQ) són determinats al grup QC/RQ o a QL, que és una unitat de vegetació propera. Els sintàxons pels que el sistema de determinació presenta una taxa d'error més elevada són els que ja hem vist que éren més conflictius: *Teucrio-Brometum subass typicum*, *Alchemillo-Festucetum*, *Euphrasio-Plantaginetum subass typicum* i diferents subassociacions d'alzinars litorals, *Viburno-Quercetum*. En aquest darrer cas cal dir que la major part de la confusió es produeix entre les diferents subassociacions de *Viburno-Quercetum*, però rarament els inventaris són assignats a altres associacions.



	AB	AD	IBL	IBT	KAT	KAA	KG	LB	TAV	TBF	TBH	TBT	TF	$N_k$
AB	21											11		32
AD1		14			1							1		16
AD2		7												7
IB			10	5	2	2		1				1		21
KA		1			28	13								42
KG							19					1		20
LB								22				4	3	29
TAV								1	13				1	15
TBF										13				13
TBH		1		1							9	4		21
TBT								2				7	3	12
TF													20	20
$N_k$	21	23	10	6	31	15	19	32	13	13	9	29	27	248

**Taula 4.1.4.A:** Matriu de confusió entre els sintàxons de *Xerobromion erecti* (en columnes) i les unitats independents de classificació de l'àrea 1 (en files). Cada casella indica el nombre d'inventaris del sintàxon de la columna que han estat classificats pel sistema a la UCI de la fila. Les caselles buides corresponen a zeros.

	AF	AP	CAT	CAG	CAF	CB	CG	EPB	EPC	EPG	EPT	GP	GT	KT	PF	PS	TAS	$N_k$
AF	11				2	1									1	1	2	18
AP	7	22																29
CA	5		26	17	9	2				2	1							62
CB*						6												6
CG	1						11				1						1	14
EPB*								5										5
EPC	2					7			10		5							24
EPG	2		1							6	1				1			11
EPT1											4							4
EPT2	2			1						2	5							10
GP	0											13						13
GT	2												12					16
KT	0													10				10
PF	1		1			1									22		2	27
PS	1															9	1	11
TAS	1			1		2					1						18	23
$N_k$	35	22	28	19	11	19	11	5	10	10	18	13	12	10	24	12	24	283

**Taula 4.1.4.B:** Matriu de confusió entre els sintàxons de *Mesobromion erecti* (en columnes) i les unitats independents de classificació (UCIs) de l'àrea 2 (en files). Cada casella indica el nombre d'inventaris del sintàxon de la columna que han estat classificats pel sistema a la UCI de la fila. Les caselles buides corresponen a zeros. L'asterisc (\*) al costat del nom d'una UCI indica que es tracta d'una unitat inactiva.

	BJ	CM	CO	MJ	OL	QCBR	QCBT	QCC	QCR	QCT	QL	QRB	QRR	QRU	RJ	RQ	$N_k$
BJ	32														1		33
CM		11		1	1												13
CO			37								2						39
MJ				23	1												24
OL1					5												5
OL2					8						1						9
QC/RQ	1		1			5	26	5	18	8	1					6	71
QL				1				1	1		51						60
QRB	4											68					72
QRR	1								1			2	8				12
QRU									2				1	22			25
RJ	3								1						9		13
$N_k$	41	11	38	24	16	5	26	6	23	8	55	70	9	22	10	12	376

**Taula 4.1.4.C:** Matriu de confusió entre els sintàxons de *Quercetea ilicis* sense *Quercenion ilicis* (en columnes) i les unitats independents de classificació (UCIs) de l'àrea 3 (en files). Cada casella indica el nombre d'inventaris del sintàxon de la columna que han estat classificats pel sistema a la UCI de la fila. Les caselles buides corresponen a valors nuls.

	AQB	AQHB	AQS	AQT	AQTL	CYB	VQAC	VQAR	VQAS	VQC	VQE	VQP	VQQP	VQS	VQV	$N_k$
AQB	31							3				2			5	41
AQHB	1	5			3										8	17
AQS			7							3				6		16
AQT				10												10
AQTL	3		1	8	15					1		2				30
CYB						11										11
VQAC							7	3							1	11
VQAR*								7				2			2	11
VQAS				1					4							5
VQC			2		1			4	3	33	7	10		4	1	65
VQE									1	2	23	1		6		33
VQP*	1							3	6	5		24	2	2	4	47
VQQP										1			12	3		16
VQS													2	56		58
VQV	1														5	6
$N_k$	37	5	10	19	19	11	7	20	14	45	30	41	16	77	26	377

**Taula 4.1.4.D:** Matriu de confusió entre els sintàxons de *Quercenion ilicis* (en columnes) i les unitats independents de classificació (UCIs) de l'àrea 4 (en files). Cada casella indica el nombre d'inventaris del sintàxon de la columna que han estat classificats pel sistema a la UCI de la fila. Les caselles buides corresponen a valors nuls. L'asterisc (\*) al costat del nom d'una UCI indica que es tracta d'una unitat inactiva.

Finalment, a la taula 4.1.5 mostrem la capacitat de resposta correcta del sistema, avaluada en les quatre àrees de classificació. Com que la resposta del sistema es una matriu difusa de pertinences, hom disposa de diversos valors de pertinença per a cada, que poden ésser

ordenats per pertinença decreixent. Per tant, és possible comptar els encerts del sistema en primera instància (comparant el sintàxon esperat amb la unitat amb un grau pertinença més elevat) o tenint en compte també la segona opció (comparant el sintàxon esperat amb les dues unitats amb pertinença més alta). A la taula 4.1.5 presentem el nombre i percentatge d'encerts del en primera opció (Enc. 1<sup>a</sup>) i considerant les dues primeres opcions (Enc. 1<sup>a</sup>/2<sup>a</sup>). A la vegada, hem desglossat en aquesta taula la capacitat d'encert del sistema per a determinar correctament inventaris utilitzats en la construcció de les UCIs (actius) de la capacitat de determinar inventaris externs (inactius).

#### Àrea 1: *Al. Xerobromion erecti*

	<i>N</i>	<i>Enc. 1<sup>a</sup></i>	%	<i>Enc. 1<sup>a</sup>/2<sup>a</sup></i>	%
Inventaris actius	119	118	99.2	118	99.2
Inventaris inactius	129	83	64.3	96	74.4
<b>Total</b>	<b>248</b>	<b>201</b>	<b>81</b>	<b>214</b>	<b>86.3</b>

#### Àrea 2: *Bromion erecti*

	<i>N</i>	<i>Enc. 1<sup>a</sup></i>	%	<i>Enc. 1<sup>a</sup>/2<sup>a</sup></i>	%
Inventaris actius	121	117	96.7	121	100
Inventaris inactius	162	99	61.1	125	77.2
<b>Total</b>	<b>283</b>	<b>216</b>	<b>80</b>	<b>246</b>	<b>86.9</b>

#### Àrea 3: *Cl. Quercetea ilicis* sense *O. Quercenion ilicis*

	<i>N</i>	<i>Enc. 1<sup>a</sup></i>	%	<i>Enc. 1<sup>a</sup>/2<sup>a</sup></i>	%
Inventaris actius	147	147	100	147	100
Inventaris inactius	229	195	85.2	215	93.9
<b>Total</b>	<b>376</b>	<b>342</b>	<b>91.0</b>	<b>362</b>	<b>96.3</b>

#### Àrea 4: *O. Quercenion ilicis*

	<i>N</i>	<i>Enc. 1<sup>a</sup></i>	%	<i>Enc. 1<sup>a</sup>/2<sup>a</sup></i>	%
Inventaris actius	121	117	96.7	118	97.5
Inventaris inactius	256	133	52	174	68
<b>Total</b>	<b>376</b>	<b>250</b>	<b>66.3</b>	<b>292</b>	<b>77.7</b>

#### Total àrees

	<i>N</i>	<i>Enc. 1<sup>a</sup></i>	%	<i>Enc. 1<sup>a</sup>/2<sup>a</sup></i>	%
Inventaris actius	508	499	98.2	504	99.2
Inventaris inactius	776	510	65.7	610	78.6
<b>Total</b>	<b>1284</b>	<b>1009</b>	<b>78.6</b>	<b>1114</b>	<b>86.8</b>

**Taula 4.1.5:** Capacitat de resposta correcta avaluada per a les quatre àrees de classificació i pel total de sintàxons coberts pel sistema de classificació. *N*: el nombre d'inventaris avaluats. **Enc. 1<sup>a</sup>**: el nombre i percentatge d'encerts considerant la primera opció de resposta. **Enc. 1<sup>a</sup>/2<sup>a</sup>**: el nombre i percentatge d'encerts considerant com a vàlides la primera o segona resposta.

Òbviament els inventaris actius de cada àrea presenten percentatges d'encert al entre el 95% i el 100% en primera instància. En canvi, la capacitat de determinació d'inventaris externs és força menor i depèn de cada àrea de classificació.

Pel que fa als resultats per àrea de classificació, l'àrea 3 és la que presenta una major capacitat de predicció (91.0% en primera instància, 96.3% en segona), però a la vegada és l'àrea en que s'ha optat per una menor capacitat de resolució. En aquesta àrea no hem considerat les diferents subassociacions de *Quercetum cocciferae*. La capacitat del sistema de determinar inventaris a les àrees de *Brometalia* és similar (80% en primera instància, 86% en segona). L'àrea de classificació amb pitjors resultats és l'àrea 4 (66.3% en primera instància, 77% en segona). La causa és la mala resposta del sistema pel que fa a les subassociacions d'alzinar litoral, que redueix sensiblement el percentatge d'encerts total. Considerem però, que la capacitat d'encert en aquesta àrea podria millorar amb un estudi més aprofundit que discernís inventaris més típics d'aquells considerats més transicionals.

Com a valoració global de la capacitat del sistema per a determinar inventaris de vegetació podem dir que, per a inventaris actius la capacitat de predicció és del 99%, mentre que per a inventaris inactius, i per tant, desconeguts, l'encert del sistema de determinació estaria al vers el 66% en primera instància i vora el 80% en segona.

### 4.1.3.3 Discussió i perspectives de futur

Hem descrit *Araucaria*, un sistema que dóna un servei remot de determinació automàtica d'inventaris de vegetació. Els detalls de la implementació del sistema de determinació es basen sobretot en els resultats dels capítols precedents. També hem proporcionat al sistema un cos de coneixement inicial: les unitats de vegetació de *Brometalia erecti* i *Quercetum ilicis*. La capacitat de resposta correcta és variable segons l'àrea de classificació, però es situa al voltant del 80% com a promitg.

Per tal que una aplicació com la descrita tingui una utilitat real, necessita incorporar al cos de coneixement del sistema nous tipus de vegetació. És important remarcar que la capacitat de resposta del sistema de determinació depèn de molts factors, que han d'ésser tinguts en compte en el procés d'inclusió de nous sintàxons. En destaquem els punts següents:

- La qualitat de les dades (inventaris) originals. Per això és important un pretractament de les dades correcte i l'aplicació de sinònimes d'anàlisi.
- Les decisions de configuració de les àrees de classificació dels exemples mostrats es sustenten sobre resultats de capítols precedents. Per a incloure noves unitats de vegetació caldria realitzar estudis d'anàlisi multivariant similars.
- L'exclusió d'inventaris intermedis en la selecció dels inventaris actius d'una àrea de classificació permet una major proporció d'unitats actives, enfront la simple definició d'unitats inactives, és a dir, no basades en mètodes de *clustering* numèrics.
- En associar un sintàxon a una unitat de classificació del sistema cal controlar que l'inventari tipus del sintàxon, si n'hi ha, formi part dels inventaris del clúster. En cas contrari, la informació proporcionada per la unitat de vegetació i el sintàxon associat farien referència a unitats de vegetació diferents.
- En certs casos, existeix compromís entre el grau de resolució sintaxonòmica i la efectivitat de la determinació. Com a exemples hom pot prendre la baixa resolució i elevada efectivitat del grup QC/RQ davant de la alta resolució i baixa eficiència de les UCI referenciades a les subassociacions de *Viburno-Quercetum ilicis*.
- En un futur, el sistema de determinació podria incloure una ponderació de variables específica de cada àrea de classificació. Tenint en compte els resultats del capítol 3.4 descartem la determinació automàtica d'aquests pesos, encara que tinguin un sentit fitosociològic. No es descarta, però, l'opció de proporcionar a l'expert la oportunitat de introduir manualment pesos que ell cregui més convenients.
- També en un futur, caldria incloure un registre local dels inventaris determinats remotament, juntament amb un registre d'usuaris no experts i experts del sistema de determinació.



## Capítol 4.2: El paquet de programes VEGANA

### 4.2.1 VEGANA

#### 4.2.1.1 Introducció

VEGANA (VEGetation edition and ANALysis, De Cáceres *et al.* 2003) és un paquet de programes destinat a la gestió y anàlisi de dades ecològiques en general i, en especial, de vegetació. Com a conjunt, pretén ésser una eina integrada de recopilació, manipulació i anàlisi d'informació ecològica. Està format per quatre programes, que poden ésser executats de manera independent (figura 4.2.1):

- Quercus*, un editor de taules de dades del tipus inventari. Permet la introducció i gestió d'inventaris d'espècies i d'altres informacions associades.
- Ginkgo*, orientat a la representació i classificació d'individus a partir de dades multivariants.
- Fagus*, un editor i gestor de citacions florístiques. Permet emmagatzemar principalment dades inèdites, bibliogràfiques o procedents de col·leccions.
- Yucca*, una eina per a realitzar representacions cartogràfiques de dades georeferenciades, principalment distribucions de tàxons, comunitats o grups d'àrees.

Els dos editors de dades, FAGUS i QUERCUS, poden importar dades del Banc de Dades de Biodiversitat de Catalunya (Font 2003, <http://biodiver.bio.ub.es/biocat/homepage.htm>). A la figura 4.2.1 es mostren les relacions de dades que existeixen entre els diferents mòduls.

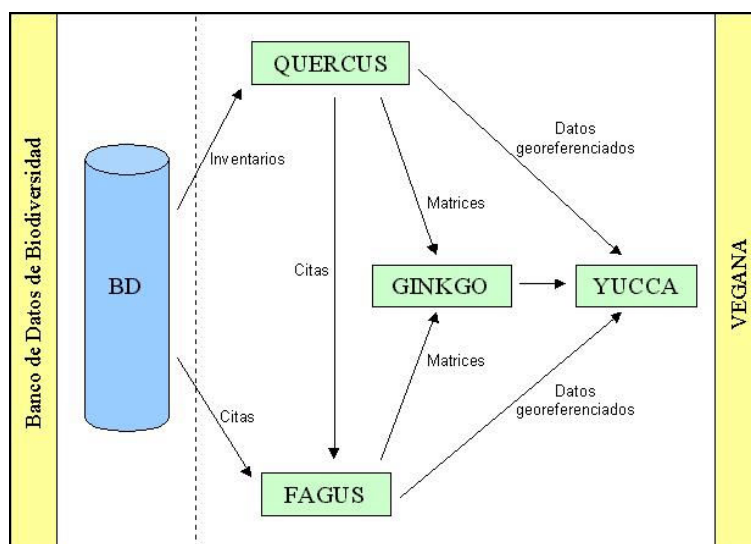


Figura 4.2.1: Esquema dels mòduls de VEGANA i les seves relacions.

#### 4.2.1.2 Características técnicas

El *software* del paquet VEGANA ha estat desenvolupat en el llenguatge de programació Java. Gràcies a això, és possible utilitzar els programes en un gran nombre de sistemes operatius (Windows, Linux, Mac, etc...). A la pràctica és possible executar els programes en qualsevol plataforma que admeti la màquina virtual de Java (*Java virtual machine*, *JVM* versió 1.4 o superior, <http://www.java.com>).

La pàgina principal del paquet VEGANA es pot trobar a: <http://biodiver.bio.ub.es/vegana>, des d'on hom pot descarregar els programes i fitxers de recursos associats. La distribució de VEGANA és gratuïta i les actualitzacions dels programes es realitzen automàticament gràcies a la tecnologia *Java Web Start*. Els requeriments de *hardware* mínims són els d'un *Pentium III* i una memòria *RAM* igual o superior a 256 *Megabytes*.

Tots els programes funcionen a través de fitxers de projecte. Un projecte conté la llista de recursos (tesaures) i fitxers de dades utilitzats en el programa, els camps específics definits per l'usuari, així com la configuració general del programa (directoris de treball, opcions visuals, idioma, ...). Tant els fitxers de dades com els fitxers de configuració dels programes es guarden en format XML (<http://www.xml.org>), pel que poden ésser visualitzats externament mitjançant navegadors o programes editors d'aquest format de dades.

En les dues properes seccions descrivim les funcions d'edició i anàlisi de la vegetació que proporcionen els programes *QUERCUS* i *GINKGO*, respectivament. Aquests són els mòduls de VEGANA que hem desenvolupat més directament relacionats amb la realització del nostre treball de doctorat.



## 4.2.2 L'editor de taules d'inventaris de vegetació *QUERCUS*

### 4.2.2.1 Introducció

*QUERCUS* ha estat concebut com una eina d'edició, manipulació i emmagatzematge d'inventaris de vegetació. Originalment, s'ha dissenyat per a facilitar la tasca de treballs fitosociològics, però també pot resultar útil per a realitzar altres tipus d'estudis ecològics basats en la manipulació de matrius de dades d'abundàncies d'espècies. Com a valors d'abundància de tàxons, *QUERCUS* admet tant els coeficients clàssics de l'escala de Braun-Blanquet com percentatges de cobertura o, simplement, el registre de la presència/absència de tàxons.

El mòdul *QUERCUS* sorgeix de l'evolució d'un altre programa ideat per a gestionar inventaris de vegetació, anomenat XTRINAU (Font, 1990; Font i Ninot, 1995), operatiu únicament sota el sistema DOS. Tot i que *QUERCUS* és bàsicament un editor de dades, permet realitzar algunes funcions semblants a les que ofereix una base de dades. Per exemple, és possible definir nous camps específics per als inventaris d'un determinat projecte i realitzar cerques senzilles sobre els valors emmagatzemats.

En els propers apartats descrivim breument les capacitats del programa. Per a una consulta més detallada vegeu el manual de *QUERCUS*:

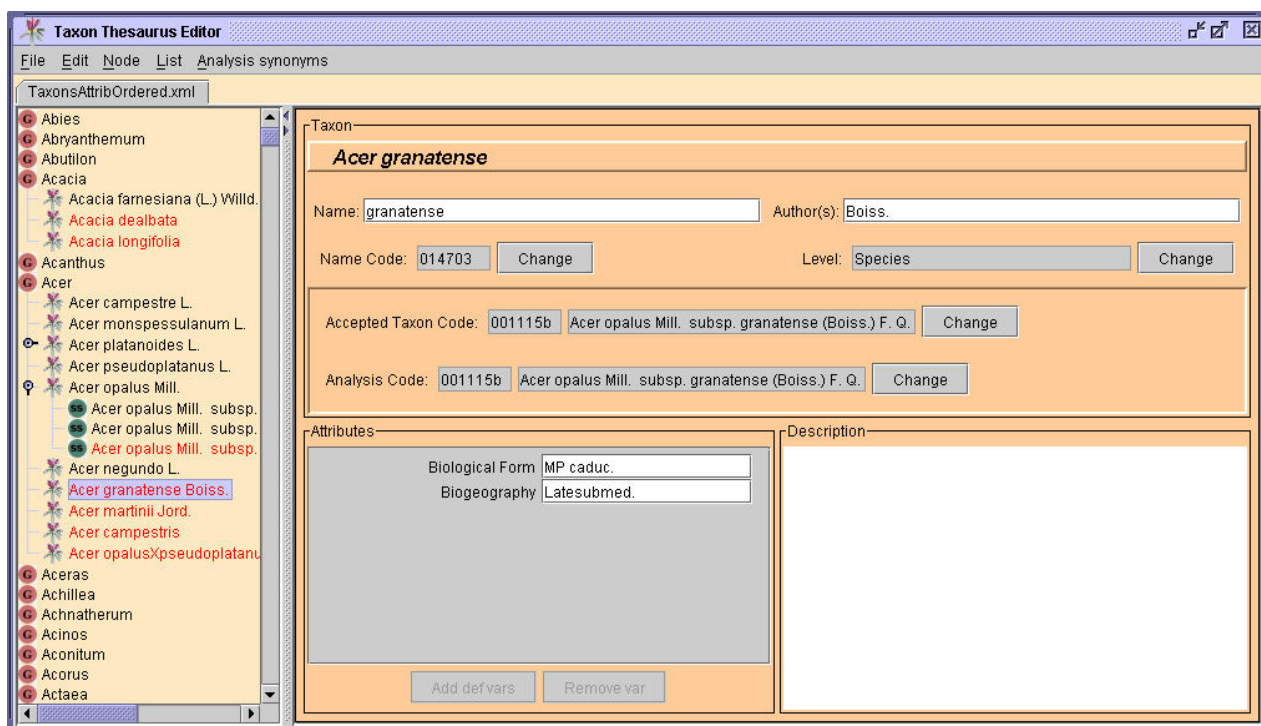
[http://biodiver.bio.ub.es/vegana/resources/help/quercus/Catala/index\\_quercus.html](http://biodiver.bio.ub.es/vegana/resources/help/quercus/Catala/index_quercus.html)

### 4.2.2.2 Tesaures

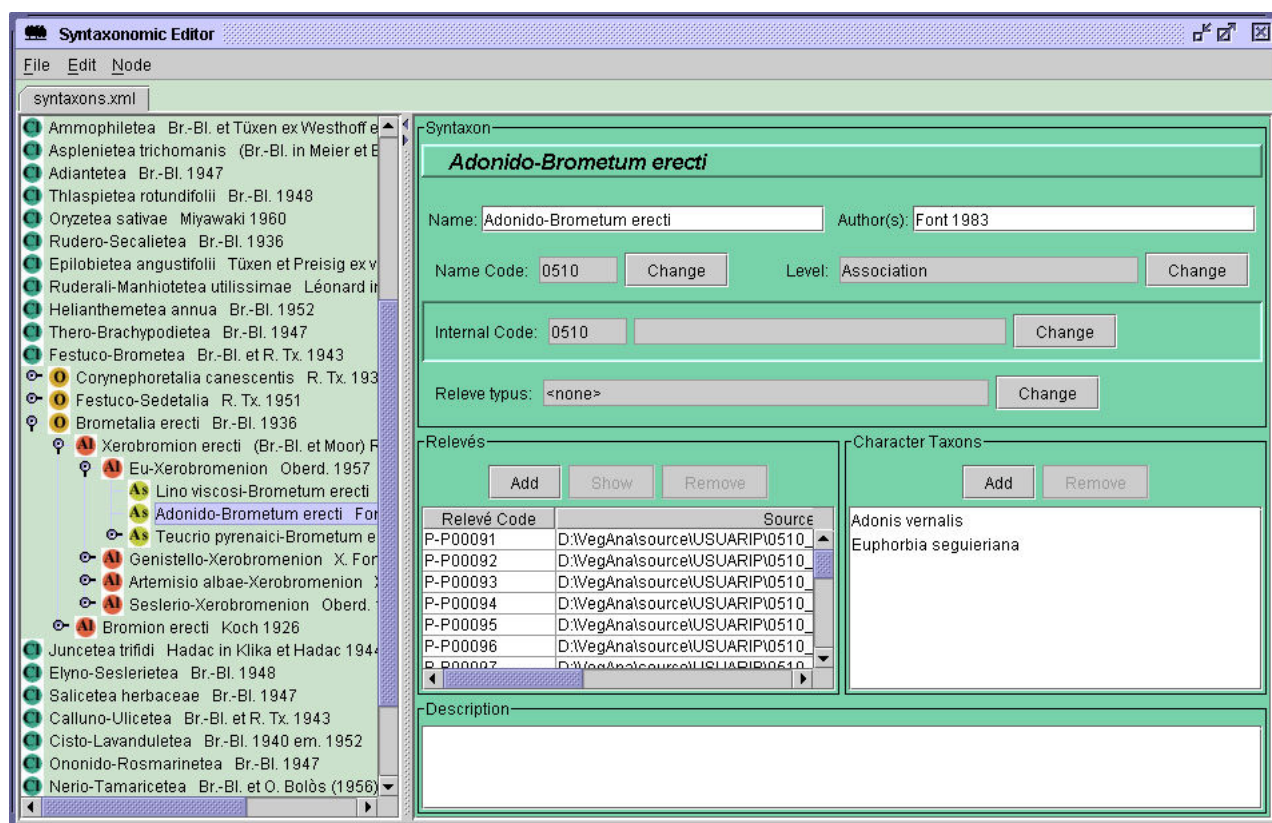
Degut a que *QUERCUS* és un editor de dades associades a tàxons, és convenient associar les dades a un fitxer de tesaure taxonòmic. L'ús d'un tesaure presenta tres avantatges: En primer lloc, els noms dels tàxons poden ésser comprovats per tal d'evitar errors tipogràfics en la fase d'introducció de dades. En segon lloc, el tesaure permet donar suport als noms sinònims, facilitant així el posterior anàlisi de dades en interpretar-se tots els noms sinònims com un únic tàxon. En tercer lloc, els tesaures poden emmagatzemar informació biològica de les espècies, cosa que fa possible la elaboració d'espectres biològics (és a dir, formes biològiques, distribució, etc.). Així doncs, per a poder utilitzar l'editor *QUERCUS* és necessari disposar d'un tesaure de tàxons. Des de la pàgina web de VEGANA es proporcionen actualment dos fitxers: un tesaure per a la flora catalana i un altre per a la flora europea. *QUERCUS* proporciona una eina d'edició de tesaures taxonòmics, que mostrem a la figura 4.2.2. Aquesta eina permet generar fitxers de tesaure propis i editar i modificar els de lliure disposició.

A banda del tesaure de tàxons, *QUERCUS* permet també utilitzar tesaures sintaxonòmics i bibliogràfics, tot i que no són imprescindibles per a la correcta utilització del programari. L'ús d'un tesaure sintaxonòmic permet associar els inventaris de vegetació a un sintàxon del tesaure.

Anàlogament a l'eina d'edició de tesaures de tàxons, és possible generar i modificar fitxers de tesaure sintaxonòmics mitjançant un editor específic (figura 4.2.3).



**Figura 4.2.2:** Aspecte de l'editor de tesaures de tàxons de *QUERCUS*. La part esquerra conté l'arbre taxonòmic, on els noms sinònims s'indiquen en vermell. A la dreta es mostren els atributs del tàxon seleccionat.



**Figura 4.2.3:** Aspecte de l'editor de tesaures de sintàxons de *QUERCUS*. La part esquerra conté l'arbre sintaxonòmic. A la dreta es mostren els atributs del sintàxon seleccionat.

### 4.2.2.3 Edició de taules d'inventaris primàries

Les taules primàries contenen inventaris encara sense elaborar des del punt de vista taxonòmic i sintaxonòmic. En relació a la definició de camps d'un inventari de vegetació, incloent la descripció de l'estructura i ecologia de la comunitat inventariada, així com la geografia de la localitat, s'han seguit les directrius proposades a Mucina *et al.* (2000). El conjunt de taules primàries associades a un projecte *QUERCUS* formen el banc d'inventaris del autor i poden incloure tan inventaris inèdits com publicats. L'editor de taules primàries de *QUERCUS* permet informatitzar i emmagatzemar inventaris de vegetació d'una manera senzilla i guiada. Presenta tres formes d'edició de les taules:

- 1) Edició de tota la taula d'entrades de tàxons alhora.
- 2) Edició de cada inventari de vegetació per separat.
- 3) Edició de la informació associada als inventari en forma de taula.

És possible importar directament a *QUERCUS* taules d'inventaris des del Banc de Dades de Biodiversitat de Catalunya (Font *et al.*, 2001). La pàgina principal del Banc de Dades es troba a <http://biodiver.bio.ub.es/biocat/homepage.html>.

### 4.2.2.4 Manipulació de taules de treball (o secundàries)

A partir dels inventaris emmagatzemats a les taules primàries s'elaboren les taules de treball o taules secundàries. Aquestes taules són les que el fitosociòleg clàssicament manipula per tal d'establir unitats de vegetació. L'editor de taules de treball (figura 4.2.4) permet realitzar aquest treball de manipulació de taules, que inclou, entre d'altres tasques: la reordenació de files i columnes, la inserció de comentaris i l'acoloriment dels inventaris per delimitar grups. L'objectiu final d'aquesta fase d'edició és preparar la taula per ésser exportada per a publicar-la o per ésser analitzada mitjançant mètodes numèrics.

Algunes operacions que es realitzen de manera automàtica a l'editor són el filtrat o suma de entrades de tàxons segons condicions específiques, com l'estrat de la comunitat en el que apareixen les plantes o la certesa de la determinació. La homogeneïtzació nomenclatural és un altre aspecte que pren importància en la preparació d'una taula per l'anàlisi i que a *QUERCUS* es realitza a l'editor de taules de treball. Hom pot distingir almenys 3 situacions bàsiques:

- 1) Eliminar els sinònims que puguin aparèixer als diferents inventaris i afegir els seus valors d'abundància al nom de tàxon vàlid.
- 2) Com que no tots els autors afinen de la mateixa manera les seves determinacions infra-específiques pot ocórrer que en una mateixa taula aparegui un tàxon determinat a nivell

d'espècie i un altre a nivell de subespècie. Llavors cal englobar les dues entrades en el tàxon de sentit més ampli, és a dir l'específic.

3) Podem trobar tàxons morfològicament molt variables i de difícil determinació.

L'editor de taules de treball permet resoldre aquests 3 casos, amb l'ajut del tesaure de tàxons. El primer cas es resol amb l'ajut del tesaure que permet reconèixer els diferents noms com a sinònims d'un tàxon. Els altres dos casos requereixen emprar un segon nivell de sinonímia, orientada a l'anàlisi, que faciliti l'automatització de tractar conjuntament dos tàxons ambdós correctes des d'un punt de vista nomenclatural i taxonòmic.

A l'editor de taules de treball és possible realitzar anàlisis numèrics simples. En primer lloc, hom pot estudiar l'estructura de la comunitat mitjançant la confecció d'espectres biològics basats en la corologia dels tàxons, el seu tipus biològic o qualsevol altre paràmetre biològic d'interès que prèviament hagi estat informatitzat al tesaure de tàxons. En segon lloc, es proporciona un senzill algorisme numèric que genera classificacions dels inventaris de la taula. L'esmentat algorisme, basat en el mètode de classificació *K-means* (MacQueen 1967), escull una mesura de dissimilaritat i decideix automàticament el nombre de grups d'inventaris en que dividir la taula.

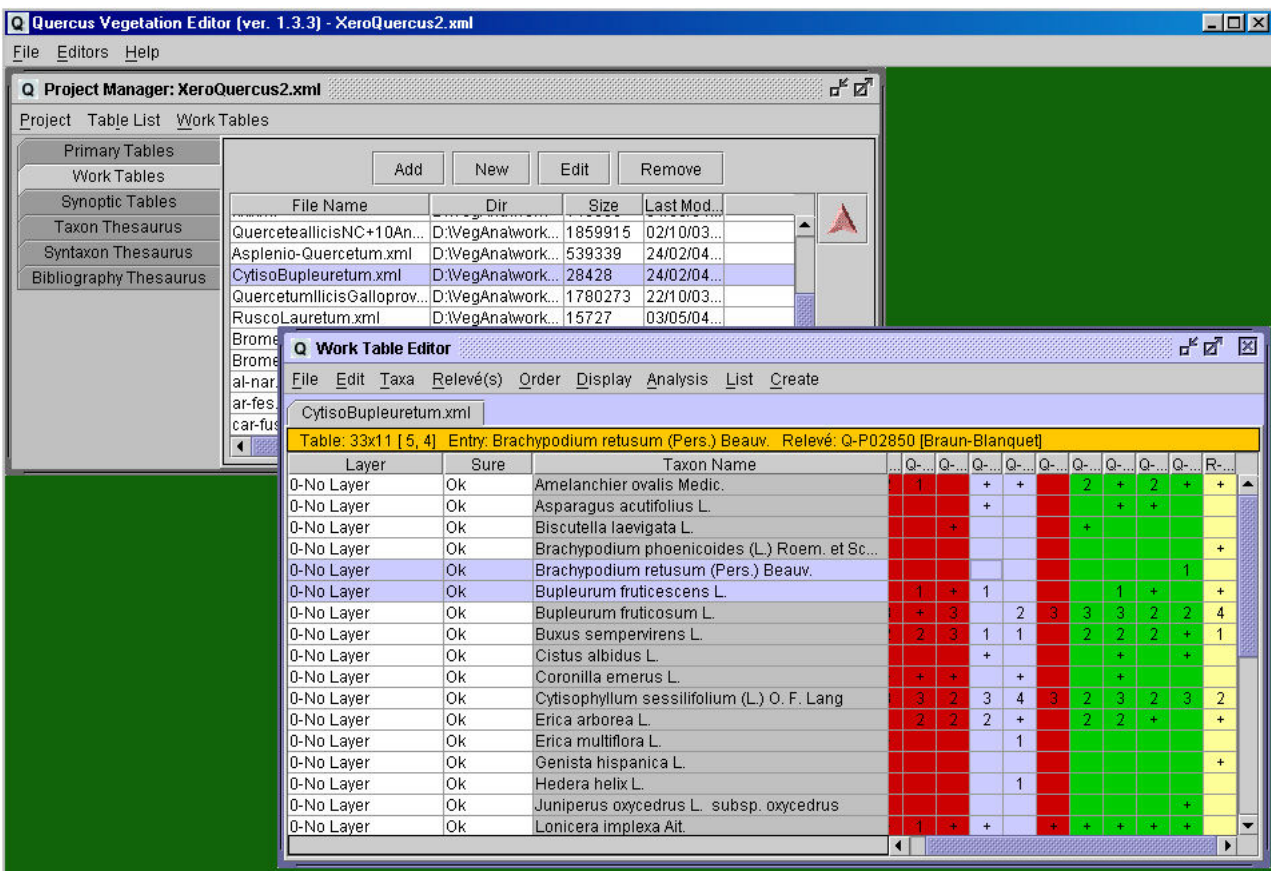


Figura 4.2.4: Aspecte general de l'editor de taules de treball de QUERCUS.

Com a una de les sortides, l'editor permet convertir una taula d'inventaris a una matriu de dades numèriques, susceptible d'ésser analitzada posteriorment amb mètodes multivariants. En el cas dels valors d'abundància mesurats en l'escala ordinal de Braun-Blanquet, és necessari transformar els valors a quantitats numèriques seguint una taula d'equivalències (van der Maarel 1979). La matriu resultant podrà ser importada a *GINKGO* o emmagatzemada en format ASCII. L'exportació de la matriu numèrica es pot complementar, en el cas de que els inventaris estiguin associats a sintàxons, amb la creació i exportació d'una partició dels inventaris en base criteris sintaxonòmics. Aquesta partició permet, juntament amb la matriu de dades, comparar la classificació obtinguda amb mètodes numèrics amb la classificació tradicional.

#### 4.2.2.5 Taules sintètiques

Una taula sintètica és essencialment com una taula d'inventaris, amb la particularitat de que cada columna conté les freqüències d'aparició dels tàxons presents en un grup d'inventaris. En altres paraules, els valors d'abundància dels "inventaris sintètics" de les taules sintètiques son valors de presència (*constancy*) expressats en percentatges o classes de percentatges.

La creació de taules sintètiques es fa únicament a partir de taules de treball. Addicionalment, si hom disposa d'un tesaure sintaxonòmic i els inventaris de la taula de treball estan associats a sintàxons, és possible crear taules sintètiques seguint la assignació sintaxonòmica dels inventaris de vegetació. Un cop creades, l'editor de taules sintètiques permet manipular les taules sintètiques. Moltes de les operacions de l'editor de taules de treball, com per exemple la manipulació de files i columnes, es troben també disponibles a l'editor de taules sintètiques.

## 4.2.3 *GINKGO*, un programa d'anàlisi multivariant basat en distàncies

### 4.2.3.1 Introducció

Les tecnologies de computació modernes haurien de facilitar l'acostament dels experts en vegetació a les tècniques d'anàlisi estadístiques. Encara que actualment existeixen molts mètodes d'anàlisi multivariant, el *software* de paquets estadístics comercials rarament inclou poc més que els mètodes més bàsics. A més, els usuaris d'aquests programes sovint estan limitats a l'ús de la Distància Euclidiana (pitagòrica), mentre que nombrosos estudis suggereixen que altres mesures de proximitat representen millor les relacions ecològiques entre inventaris de vegetació (p.ex. Hajdu 1981, Faith et al. 1987). Els usuaris experts poden evitar aquesta limitació transformant *a priori* les dades (see Legendre & Gallagher 2001) o aplicant tècniques de *scaling* (Gower 1966; Kruskal 1964a, 1964b). No obstant, molts usuaris no experts encara accepten les limitacions del *software* disponible. Afortunadament, existeixen força programes orientats específicament al camp de l'ecologia de la vegetació. Algunes d'aquestes eines ofereixen només alguns mètodes numèrics. Alguns exemples són TWINSpan (Hill et al. 1974, Hill 1979), CANOCO (ter Braak & Šmilauer 1988) o JUICE (Tichý 2002). Altres eines, com MULVA-5 (Wildi & Orłóci 1996) són molt complertes en termes del nombre de mètodes disponibles però els falta una interfície gràfica d'usuari senzilla. Les opcions de *software* que potser són més complertes en ambdós sentits són SYNTAX (Podani 1993, 1994), ADE-4 (Thioulouse et al. 1997), PC-ORD (McCune & Mefford 1999), i el *R-package* (Casgrain & Legendre 2001). No obstant, tots aquests programes són específics d'una plataforma i sistema operatiu. Veient aquestes mancances hem cregut oportú el desenvolupament d'un nou paquet estadístic orientat a l'anàlisi de vegetació, multiplataforma i de fàcil utilització.

L'objectiu d'aquesta secció és presentar el programa *GINKGO*. El programa s'orienta principalment cap als mètodes d'anàlisi multivariant de matrius de dades ecològiques basats en distàncies. Aquesta orientació ha estat presa en dos sentits: En primer lloc, proporcionant moltes mesures de similitud i dissimilitud, algunes de les quals tenen propietats desitjables per a expressar relacions ecològiques o taxonòmiques. En segon lloc, implementant mètodes d'ordenació i classificació aplicables a matrius simètriques de dissimilitud. Aquests mètodes presenten l'avantatge de conservar, durant el procés d'analitzar l'estructura de les dades, l'espai de proximitats entre inventaris que proporciona la mesura de dissimilitud escollida. En les properes seccions, descrivim les característiques principals del programa. En primer lloc presentem la interfície gràfica d'usuari i el funcionament general del programa. En segon lloc s'enumeren els formats d'emmagatzematge i intercanvi de dades. A continuació, citem els mètodes multivariants d'ordenació i classificació que el programa presenta, explicant breument aquells mètodes basats en distàncies que considerem nous per a l'anàlisi de vegetació.



Finalment ens dediquem als mètodes de comparació i avaluació d'estructures de classificació. Per a una consulta més detallada vegeu la pàgina web del manual de *GINKGO*:

[http://biodiver.bio.ub.es/vegana/resources/help/ginkgo/index\\_ginkgo.html](http://biodiver.bio.ub.es/vegana/resources/help/ginkgo/index_ginkgo.html)

#### 4.2.3.2 Generalitats

*GINKGO* és un programa que intenta facilitar l'ús de tècniques d'anàlisi multivariant entre usuaris no experts en estadística. Ha estat desenvolupat en el camp de la ciència de la vegetació, però pot ésser emprat en altres disciplines científiques (p.ex. sociologia o psicologia). El programa presenta una interfície d'usuari que comprèn tres finestres principals: un editor de dades, un gestor d'anàlisis, que emmagatzema tots els anàlisis que l'usuari realitza, i un editor de gràfics (veure figura 4.2.5). La interfície com a conjunt proporciona un marc de treball integrat que permet a l'usuari explorar, pas a pas, les seves dades.

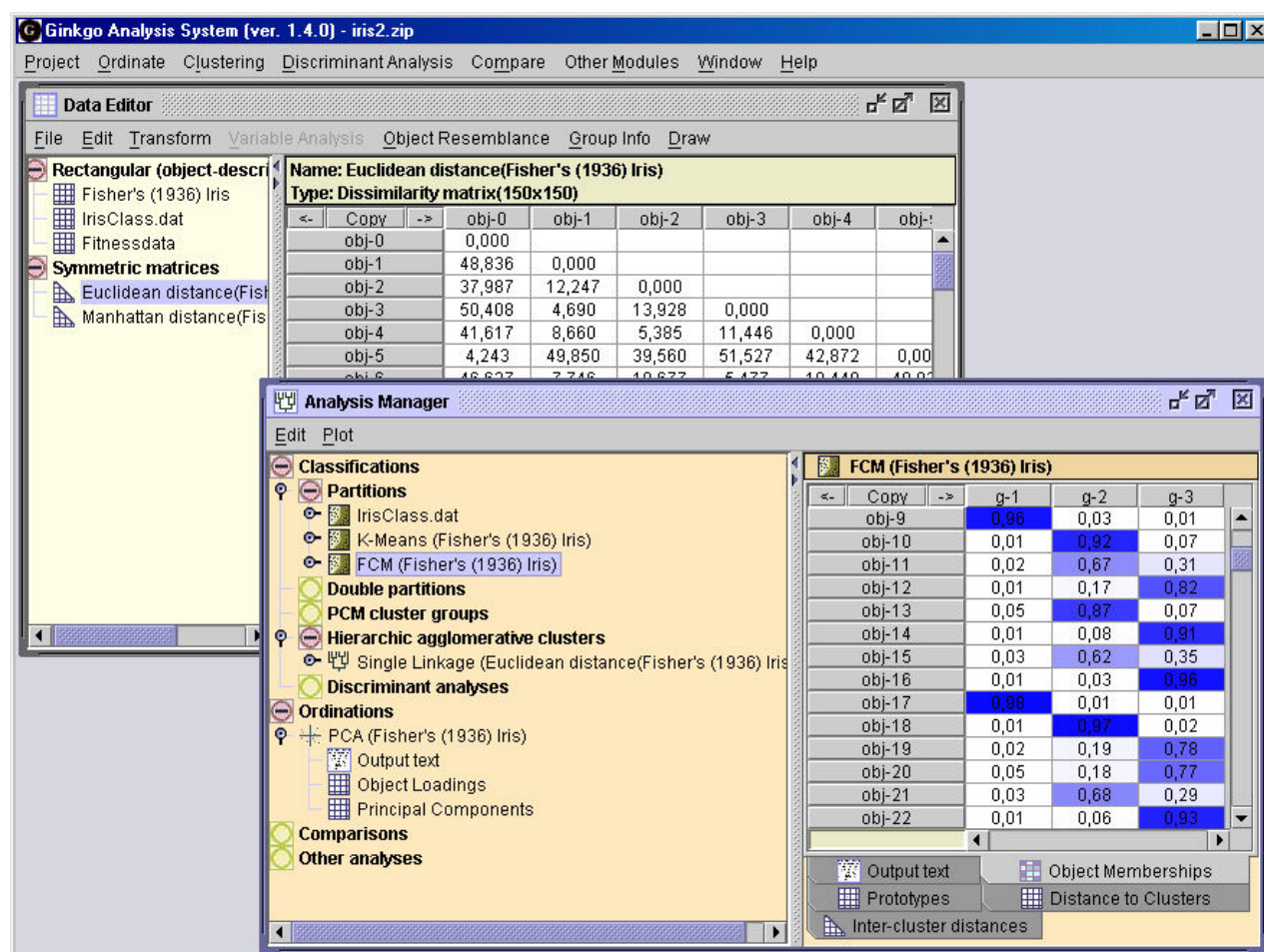


Figura 4.2.5: Aspecte general del programa *GINKGO*, amb les finestres de l'editor de dades i el gestor d'anàlisi.

En una activitat de treball normal, els passos implicats en l'estudi de l'estructura d'una matriu de dades de vegetació són: En primer lloc hom pot aplicar transformacions escalars o vectorials. En segon lloc, si hom desitja emprar mètodes multivariants basats en matrius de distància, cal calcular la semblança entre inventaris o espècies, escollint una mesura de proximitat entre les

diferents similaritats i dissimilaritats. En el cas de les matrius de similaritat és necessari una transformació *a posteriori* per a obtenir matrius de distàncies. A continuació, hom executa típicament mètodes d'ordenació i de classificació de dades. Els resultats que s'obtenen poden ésser copiats de nou a l'editor de dades per a posteriors anàlisis. En conjunt, el programa permet dur a terme anàlisis multivariants de complexitat moderada. El projecte de treball de *GINKGO*, incloent dades i anàlisis acumulats, es pot guardar en un únic fitxer comprimit (.zip). Això permet realitzar l'estudi de l'estructura de les dades en sessions diferents.

#### 4.2.2.3 Edició i intercanvi de dades

L'editor de dades de *GINKGO* permet importar, crear i editar matrius de dades multivariants. S'hi diferencien dos tipus de matrius (veure figura 4.2.5): matrius rectangulars (objecte-descriptor) i matrius simètriques (objecte-objecte o descriptor-descriptor). Matrius d'ambdós tipus poden ésser creades o importades. Per a matrius rectangulars es proporcionen diferents opcions d'importació de text ASCII pla: fitxers de text amb diferents caràcters separadors de valors, fitxers de espècies en format Cornell condensat o fitxers de variables ambientals de CANOCO. L'exportació de matrius de dades es pot realitzar en els mateixos formats de fitxers ASCII o a través del portapapers del sistema. A més, és possible importar taules d'inventaris de vegetació en format *xml* creats amb *QUERCUS* o provinents del banc de dades de biodiversitat.

Com és freqüent en altres programes d'estadística, hom pot calcular a *GINKGO* estadístics de descripció univariant de les dades, així com generar matrius de covariància i correlació de variables. Altres utilitats corrents disponibles són les transformacions escalar i vectorial, la estandardització i la transposició de matrius.

#### 4.2.2.4 Mètodes d'ordenació

Els mètodes d'ordenació o de reducció de la dimensionalitat inclosos a *GINKGO* es troben a la llista de la taula 4.2.1. Els mètodes d'ordenació no-canònica parteixen d'una sola matriu de dades (rectangular o simètrica) i proven de representar la informació continguda en unes poques dimensions. Els mètodes canònics persegueixen la representació de les dades partint de la combinació de dues matrius. La manera com es combinen difereix, evidentment, en cada mètode. D'una banda, l'anàlisi de la redundància (RDA, van den Wollenberg 1977) i l'anàlisi canònica de correspondències (CCA, ter Braak 1995) utilitzen una de les matrius com a matriu de variables explicatives i l'altre com a matriu resposta. En canvi, *related multidimensional scaling* (RMDS, Cuadras & Fortiana 1998) és una anàlisi canònica simètrica que, com a tal, tracta les dues matrius d'entrada de la mateixa manera (és a dir, són intercanviables). Aquest mètode d'ordenació pot ésser considerat com la contrapartida canònica simètrica de l'Anàlisi de Coordenades Principals (PCoA/MDS, Gower 1966). RMDS combina la informació de proximitat entre objectes continguda en les dues matrius originals i en derivar una representació en un



espai Euclidià. Aquest mètode presenta l'avantatge de prendre en consideració i descartar aquella informació redundant de les matrius originals.

Mètodes no-canònics	Mètodes canònics
<i>Principal component analysis</i> (PCA)	<i>Redundancy analysis</i> (RDA, van den Wollenberg 1977)
<i>Correspondence analysis</i> (CA, Hill 1973)	<i>Canonical correspondence analysis</i> (CCA, ter Braak 1995)
<i>Metric multidimensional scaling/Principal Coordinates Analysis</i> (MDS/PCoA, Gower 1966)	<i>Related metric scaling</i> (RMDS, Cuadras & Fortiana 1998)
<i>Non-metric multidimensional scaling</i> (NMDS, Kruskal 1964a, 1964b)	

**Taula 4.2.1:** Mètodes d'ordenació disponibles a GINKGO. Els mètodes de la columna de la dreta són la contrapartida canònica dels mètodes no canònics de l'esquerra.

#### 4.2.2.5 Mètodes de classificació

Un dels aspectes més problemàtics en estudis ecològics que analitzen dades mitjançant mètodes multivariants és la determinació de l'estructura grups present a les dades. En l'apartat de mètodes de *clustering* GINKGO permet aplicar 3 models de clúster diferents (Taula 4.2.2, columna de l'esquerra). En primer lloc trobem el model jeràrquic clàssic. Els mètodes de *clustering* jeràrquics disponibles en el programa són els mètodes aglomeratius ja molt coneguts, com *single linkage*, *complete linkage*, *UPGMA*, *WPGMA*, *UPGMC*, *WPGMC*, *Ward's minimum variance*, and *the  $\beta$ -Flexible clustering*. Tots ells s'executen a GINKGO a partir d'una matriu simètrica de similaritats o dissimilaritats.

El segon model de *clustering* el conformen les particions d'objectes, tant *crisp* com *fuzzy*. GINKGO incorpora dos mètodes de *clustering* partitiu: El ja clàssic *K-means* (MacQueen 1967), i *Fuzzy C-Means* (FCM, Bezdek 1981). FCM generalitza *K-means* a la lògica difusa, i fou aplicat a l'anàlisi de la vegetació a partir dels treballs de Marsili-Libell (1989) i Equihua (1990). És important ressaltar que la implementació d'aquests mètodes a GINKGO permet la seva execució a partir de matrius dissimilaritat, apart del seu ús normal a partir de matrius rectangulars (Oliva et al., 2001). Aquesta opció es troba rarament en altres eines de software d'anàlisi multivariant. Per a executar els *K-means* i FCM sobre matrius de similaritat, aquestes han d'ésser, òbviament transformades a dissimilaritats prèviament. Un altre mètode de *clustering* que genera particions és *REBLOCK* (Podani & Feoli 1991). *REBLOCK* produeix a la vegada una partició dels inventaris i una partició dels tàxons, configurant una estructura combinada de blocs, similar al treball fitosociològic d'ordenació de taules d'inventaris. La implementació que GINKGO proporciona d'aquest mètode maximitza la contingència entre blocs, mitjançant una funció basada en l'estadístic  $\chi^2$ .

El darrer model de *clustering* és un model de clústers independents, això equival a dir que cada clúster es determina independentment dels altres. La utilitat d'aquest model rau en la detecció de regions d'alta densitat d'objectes (p.ex. inventaris), concepte similar al de 'node' de vegetació (Wildi 1979). La implementació algorísmica d'aquest model a *GINKGO* la proporciona l'algorisme *Possibilistic C-Means (PCM)*, Krishnapuram & Keller 1993, 1996). El model possibilístic sorgeix de la relaxació del concepte de partició *fuzzy* (veure capítol 3.3). En una partició de *FCM* cada objecte està limitat a que la suma dels seus valors de pertinença sigui 1. Els graus de pertinença es calculen a partir de la comparació de les distàncies entre l'objecte i cada un dels centroids dels clústers. Aquests graus de pertinença són, doncs, *relatius*. En canvi, els objectes classificats per *PCM* no tenen la limitació esmentada. El seu grau de pertinença pot ésser considerat *absolut*, i s'anomena tipicalitat. La tipicalitat que un objecte mostra per un clúster s'obté de la comparació de la distància entre l'objecte i el centroid del clúster amb una distància de referència, que és un paràmetre del mètode de *clustering*. La implementació de *PCM* del programa *GINKGO* s'anomena *meta-PCM*, perquè inclou dues modificacions de l'algorisme original (veure capítol 3.2): 1) Una correcció de la funció de tipicalitat per a mesures de dissimilaritat acotades, i 2) Un algorisme que suggereix la millor distància de referència per a cada clúster de *PCM*. De manera anàloga a *K-means* i *FCM*, l'algorisme *meta-PCM* pot ésser executat a partir de matrius de dades rectangulars o matrius simètriques de dissimilaritat. Com a sortida genera clústers difusos esfèrics, que en alguns casos poden incloure's els uns als altres indicant diferents nivells de densitat anidats.

Quan hom disposa d'una partició dels objectes *a priori*, és possible realitzar una anàlisi discriminant que permeti classificar noves observacions. *GINKGO* proporciona tres funcions d'anàlisi discriminant diferents (taula 4.2.2, columna de la dreta): L'anàlisi discriminant lineal canònica (CDA), l'anàlisi discriminant quadràtica (QDA) i l'anàlisi discriminant basada en distàncies (Cuadras *et al.* 1997). L'anàlisi discriminant basat en distàncies utilitza matrius d'entrada de dissimilaritats. La funció discriminant es construeix de manera similar als mètodes de *clustering* partitius. És important ressaltar que tots els mètodes d'anàlisi discriminant esmentats poden utilitzar particions *fuzzy* com a particions d'entrenament.

Mètodes de <i>clustering</i>	Mètodes d'anàlisi discriminant
Mètodes jeràrquics aglomeratius	<i>Canonical linear discriminant analysis (CDA)</i> .
<i>K-means</i> (MacQueen, 1967)	<i>Quadratic discriminant analysis</i>
<i>Fuzzy C-means</i> (Bezdek 1981)	<i>Distance-based discriminant analysis</i> (Cuadras et al., 1997).
<i>REBLOCK</i> (Podani & Feoli 1991)	
<i>Possibilistic C-means</i> (Krishnapuram & Keller 1993)	

**Taula 4.2.2:** Mètodes de *clustering* i anàlisi discriminant disponibles a *GINKGO*.

Els resultats dels mètodes d'ordenació es poden visualitzar en diagrames de dispersió de dues i tres dimensions. Els diagrames que es poden obtenir inclouen, per a alguns mètodes d'ordenació, *biplots* i diagrames de Shepard (1962). Hom pot diferenciar en aquests diagrames els objectes, mitjançant la utilització de matrius de classificació, *fuzzy* o *crisp*. D'aquesta manera és possible interpretar alhora els patrons revelats per les estratègies d'anàlisi de la classificació i la ordenació, que són complementàries.

La finestra de l'editor de gràfics admet algunes accions d'edició dels diagrames que s'hi mostren, com ara canviar opcions dels eixos, canviar títols o redimensionar els gràfics. També incorpora la possibilitat d'imprimir els gràfics, o exportar-los en un fitxer d'imatge en format *jpg*, *gif* o *png*.

#### 4.2.2.6 Avaluació i comparació de classificacions

La validació d'estructures de *clustering* pot realitzar-se emprant criteris *interns*, *externs*, o *relatius* (Jain & Dubes 1988: p. 161). Els criteris interns estimen l'ajust entre la estructura trobada i les dades originals. Poden servir per a decidir, per exemple, el nombre de grups a cercar en unes dades. Els criteris externs mesuren la validesa de la classificació comparant-la amb una estructura externa, presa *a priori* com a vàlida. Finalment, els criteris relatius comparen diferents classificacions entre elles.

*GINKGO* incorpora dos criteris interns per particions: L'estadístic pseudo-F de Calinski-Harabasz (1974), i l'estadístic no paramètric *silhouettes* (Rousseuw 1987), les quals poden emprar-se per avaluar la bondat de la classificació d'objectes individuals. Per a avaluar particions amb criteris externs i/o relatius, el programa proporciona l'índex de Rand corregit (Rand, 1971; Hubert & Arabie, 1985) i l'índex de Fowlkes-Mallows (1983). Aquests índexs estimen el nivell d'acord entre dues particions *crisp*. Per tal de comparar particions *fuzzy*, aquestes han d'ésser normalment 'defuzzificades'. Com a alternativa, *GINKGO* proposa l'ús d'una modificació de l'índex de Rand que permet comparar particions difuses directament. Una altra manera de comparar-les és mitjançant l'índex MINDMT de Podani (1990). Les comparacions entre classificacions jeràrquiques (dendrogrames) poden realitzar-se generant particions a diferents nivells de proximitat i establint a continuació el grau d'acord amb els índexs esmentats (Fowlkes & Mallows 1983). Al mateix temps, aquesta estratègia de comparació permet comparar particions amb dendrogrames. La realització de comparacions a *GINKGO* no estan restringides als resultats obtinguts al mateix programa. Efectivament, és possible importar classificacions externes al gestor d'anàlisi, a partir de fitxers ASCII plans o llegint els resultats de TWINSpan que genera el programa PC-ORD (McCune & Mefford 1999). Aquesta darrera opció, d'utilitat en estudis sintaxonòmics, importa la doble partició d'inventaris i espècies que proporciona TWINSpan.

