Acta Bot. Gallica, **156** *(4), 607-624, 2009.*

Determination of the nutritional state of soils by means of the phytosociological method and different statistical techniques (Bayesian statistics and decision trees) in Spain

by Ana Cano-Ortiz(1), Carlos J. Pinto Gomes(2), Francisco J. Esteban(3) and Eusebio Cano Carmona(4)

(1) Dpt. of Agricultural Engineering. TRAGSATEC, Madrid (Spain); acao@tragsatec.es
(2) Dpt. of Ecology. University of Évora, P-07000 Evora; cpgomes@uevora.pt
(3) Dpt. of Experimental Biology. University of Jaén, E-23071 Jaén
(4) Dpt. of Animal Biology, Plant Biology and Ecology (Botany), University of Jaén, E-23071 Jaén; ecano@ujaen.es

Abstract.- The paper deals with the nitrophilous *Hordeum leporinum* pasturelands peculiar to the olive groves of Jaén, Spain. The study first selects physiognomically homogeneous land plots and subsequently samples are taken where the relevant species is dominant. Simultaneously, an edaphic sample is taken and the different edaphic parameters are analysed. By means of phytosociological and statistical methods (cluster analysis) and co-location networks, we describe a new association, namely *Anacyclo clavati-Hordeetum leporinum.* With the aid of advanced statistical techniques and Bayesian networks, we carry out different analyses in order to find out plant-soil correlations, in particular, for *Filago lutescens* and *Hordeum leporinum.* The paper also provides a decision tree for *Hordeum leporinum.* In this decision tree the values of the different edaphic parameters are checked against the values of the abundance-dominance index.

Key words : pasturelands - association - statistics - networks - artificial intelligence.

Résumé.- Nous avons réalisé une étude des herbages nitrophiles à *Hordeum leporinum* dans les olivettes de Jaén, Espagne. Nous avons choisi des portions de terrain homogènes physionomiquement et avons fait un relevé où l'espèce caractéristique domine. Parallèlement, nous avons fait un échantillonnage du sol et avons analysé les différentes paramètres édaphiques. À l'aide de méthodes phytosociologique et statistiques (cluster) et des grilles d'emplacement, nous décrivons la nouvelle association *Anacyclo clavati-Hordeetum leporinum.* À l'aide de techniques statistiques avancées, nous avons réalisé différentes analyses pour établir une corrélation plante-sol, dont les réseaux bayésiens pour *Filago lutescens* et *Hordeum leporinum.* Nous avons élaboré un arbre de décision pour *Hordeum leporinum,* où on établit un rapport entre les valeurs des différentes paramètres édaphiques et la valeur du taux d'abondance-dominance.

I. INTRODUCTION

regions of the Spanish territory, olive trees are the main crop. As a result of spe-nate conditions, olive trees are exploited in extensive areas of monoculture. ; a large portion of the surface occupied by olive tree groves is still exploited by f socioculturally outdated farming techniques and, consequently, many farms are ductive.

result of the ever-increasing soil erosion, aquifer pollution, loss of biodiversity, of alien floristic elements, pest population unbalances and destruction of ecosys-is extremely urgent to adopt, in larger areas of the territory, an environmentally agriculture which can be simultaneously productive and profitable for farmers. reason, our aim here is to study the vegetal covers (*Hordeum leporinum* commu-; potential « edaphic bioindicators ».

haustive study of these communities and a deep revision of the multiple edaphic carried out on the same sites by Knoepp *et al.* (2000) have provided useful sta-iformation to correlate the occurrence of taxa o syntaxa to certain soil features (or n elements growing on the soil) peculiar to olive tree groves. In this manner we gn treatment and management programmes to make natural vegetal covers profi-

ommunities under study are typically mediterranean, subnitrophilous or nitrophi-sslands developing in spring time. This therophytic and ruderal vegetation thrives loned farmlands and viary areas. Phenologically developing in spring time or early these plants also irradiate to submediterranean and central European territories.

of the communities under study is the *Bromo scoparii-Hordeetum leporinum* asso-Rivas Martínez, 1978). The community tends to grow in the meso- and suprame-ean thermotypes, presents a therophytic character and a vernal phenology and was d for the continentalised environments of the centre of Spain. Towards less nitri-dry soils, the community makes catenal contact with the *Taeniathero-Aegilopion tae* pasturelands; if the moisture level is higher, with the vallicares of *Agrostion cae* and, towards more nitrified areas, with the *Malvenion neglectae* grasslands.

ritories near to the Guadalquivir river (Hispalensean Sector) and in the Subbetic here is a *Hordeum leporinum* pastureland which develops on basic substrates with es higher than 8. These pasturelands frequently present basophilous species mis-he *Bromo scoparii-Hordeetum leporinum* community described for the continen-reas of the Meseta (Rivas-Martínez, 1978). The optimal distribution area of this n is located in the mesomediterranean, dry-subhumid belt of the Betic territories, irly in the area of the series *Paeonio-Querco rotundifoliae* s. and *Viburno tini-fagineae* s.

II. MATERIALS AND METHODS

ly identified the communities under study and determined their location and rela-to the olive tree groves. By so doing, we delimited the study area, namely, the NW aén province. We then concentrated in the countryside of the Guadalquivir river nd sierra Morena (Jaén), an area extensively occupied by olive tree groves and, to extent, by cereal, cotton and sunflower crops.

The area under study is encompassed within the Betic Province, Hispalensean Sector. This Sector is located in the Guadalquivir river valley and extends over a large territory with a great variety of soil materials. In the deeper areas accumulated marls produce the so-called bujeos (muddy soils) when organic matter is abundant. In the peripheral area surrounding the depression there are also loose, sandy or muddy materials at the foot of the Betic ranges. Quaternary alluvial materials tend to dominate with silts and Miocene marls. Meanwhile, in the northern section of the territory, some Triassic outcrops can also be found.

Notwithstanding the important role that physical factors, such as climate, soil and geography, have played in the configuration of this or that vegetal ecosystem, it is man action that dramatically determines the current landscape profile. According to the impact of this man-made alteration, we must distinguish between natural, rural, urban and industrial ecosystems.

Since the study area is far from the coastline, summers tend to be very dry and hot, whereas winters are cold due to the continentality of the nearby Meseta.

Seven bioclimates have been defined within the mediterranean macroclimate. This paper concentrates on an area under pluviseasonal-oceanic conditions, with $Ic \leq 21$ and $Io > 2.0$ (Rivas-Martínez, 1996).

The thermotype of the area under study ranges from upper thermomediterranean to lower mesomediterranean. The dominant ombrotype is dry and subhumid.

For the edaphic analyses we took soil samples, namely 1 kg at different depths and following the radicular system of the dominant species. These soils were later identified with the same code used in the phytosociological relevés made in that particular UTM following the phytosociological method (Braun-Blanquet, 1979). The edaphic analyses were carried out by the Laboratorio Agroalimentario of Granada (Atarfe) following the analysis protocol stated in the BOE 246, order of 17/09/1981. The edaphic parameters (*i.e.* soil features) monitored were (the abbreviations and corresponding measurement units are given in brackets): exchange calcium (Ca, meq/100 g), cationic exchange capacity (CEC, meq/100 g), carbonates (%), assimilable phosphorus (P, p.p.m), exchange magnesium (Mg, meq/100 g), oxidisable organic matter (OMM, %), total nitrogen (N, %), pH 1/2.5, exchange potassium (P, meq/100 g), pF 1/3 atmosphere (%), pF 15 atmosphere (%), clay texture (Txclay, %), silt texture (Txsilt, %), sand texture (Txsand, %), sieve 2 mm (%), previous salinity test (mmhos/cm).

In order to determine the different taxa for the floristic analysis, we used as reference works those of Castroviejo *et al.* (2001), Valdés *et al.* (1987), Tutín *et al.* (1964-93) and *Flora de Portugal* (Pereira Coutinho, 1939).

The species belonging, for certain statistical analyses, to restricted groups were selected according to particular criteria. Thus, the species conferring the name and physiognomical character to a community must all be characteristic species of that community and occur in at least 40% of the phytosociological relevés made for that community. If those species are not present in at least 40% of the relevés, those species must then be ecologically restricted species.

In order to first check the qualitative and quantitative relationships between land plots, we implemented statistical techniques and methods. For this purpose we carried out multiple correlation analyses to obtain the correlation coeficient for each pair of land plots and the *p*-values (significance values) for the correlation. By means of the *Matlab* 7.1 programme we generated a matrix of correlation coefficients out of the matrix of Van der Maaler indexes, with rows corresponding to species (record), and columns corresponding to land plots (variable). Species were arranged in ascending alphabetical order while land plots were arranged according to releve number and grouped by associations. The matrix of correlation coefficients was in turn obtained from a covariance matrix (C) by means of the following formula: C(i,j)/SQRT(C(i,i)*C(j,j)).

We also obtained a matrix of *p*-values to put to test the hypothesis of the absence of correlation. Each *p*-value represents the probability of obtaining such a high correlation, as if the recorded value were a random value with a true correlation being equal to null. If the p-value is lower than 0.05, the correlation between a pair of land plots is deemed to be significant. The *p*-value was computed by transforming the correlation matrix into a statistical value *t* with *n*-2 degrees of freedom, *n* representing the number of species. Confidence limits are based on an asymptotic normal distribution of 0.5*log((1+R)/(1-R)), with a variance approximately equal to 1/(*n*-3). When data follow a multivariate normal distribution pattern, these limits are reliable for large samples.

In addition to the phytosociological correlation analyses, we also used classical statistical methods to obtain the following descriptive parameters for each of the edaphic features of the association: mean, median, variance, standard deviation, standard error, minimum, maximum, range, first quartile, second quartile and coefficient of variation.

A correct implementation of classical statistical methods requires very strict conditions, such as a fixed probabilistic model (usually a normal distribution) in which only one or two parameters can remain undetermined (the mean and/or the variance). Such a restriction poses a serious problem because the normally used probabilistic models are hardly suitable for the random phenomenon being recorded. For this reason the results obtained under such conditions (normality and homoscedasticity) are no longer reliable.

Our approach makes use of robust statistical methods to analyse and compare the edaphic features of each association. Since, as far the robustness is concerned, the fundamental problem of the sample mean is that it is highly sensitive to the occurrence of extreme sample values, as robust mean estimator we used the *sample Winsorized mean* (in the tables, abbreviated as *media-win*). As robust median estimator we used the *Huber M-estimator of location* (in the tables, abbreviated as *mediana-est*), since if the model presents anomalous data, no record will be altered and such an estimator will coincide with the sample mean (which is, actually, the most reliable estimator). Finally, we also calculated the sampling error of the estimator (*nqse*) and the sampling error of the sample median (*stse*).

To obtain robust estimators we used the R^{mo} software, which is the generic name referred to the robust statistical free software package R (The R Projet for Statistical Computing; *http://www.r-project.org/*; García Pérez, 2005).

The study concentrates on the phytosociological method and implements co-location networks, Bayesian statistics and the technique of decision trees.

We used Bayesian networks to graphically represent a set of related uncertainties. These networks are based on the theory of probability (Bayes' theorem) and the theory of graphs models to represent reality. We have used Bayesian networks for plant-soil relationships. This probabilistic approach presents a qualitative and a quantitative dimension. From the qualitative point of view, Bayesian networks can be interpreted as a graphical structure (a graph or network) representing a set of dependance relationships between variables. The kind of networks so obtained are called *directed acyclic graphs*. Each node represents a variable and arcs the actual relationship between the nodes involved. This sort of modelling has the advantage of allowing the coding of conditional dependance (and independance) relationships between variables, which makes the interpretation an calculuses much easier.

In the quantitative approach the uncertainty associated to each variable of a Bayesian network is treated with the theory of probability. Thus, each variable of a model can present a limited number of states (or levels) and each of these states is associated to a value, which corresponds to the probability of that state. When we have some previous evidence about any of the variables involved in the model, these probabilities can be accordingly modified. In that case, *i.e.* when we actually know the value which a given variable of the model has, that information propagates through the network and the values associated to other variables are then recalculated by means of the Bayes' theorem as basic operator. This is known as *evidence propagation*. Thanks to the codified principles of conditional dependance and independance used for the graphical structure and to the algorithms developed to operate with them, this calculuses can be carried out in a much shorter time than it would take if the whole set of variables involved in the model were to be dealt with.

Nowadays there are several programmes (software) available to model Bayesian networks. We have used the programme *Bayesia Lab* 3.1 (*http://www.bayesia.com*). *Naive Bayes* was the classifier used to characterize each node, since it relies on the assumption of a heavy independance between the characteristics in the classification process, *i.e.* in the ascription of a class mark to each of the instances (in our case, plant species) described for a series of features (in our model, edaphic parameters). In order to obtain a value indicating us the potential information provided by the knowledge of each soil feature according to the species growing on it, we used the Kullback-Leibler divergence value. This value indicates the information gain associated to a given node (edaphic feature) for the knowledge of a target node (the relevant species) (Cano-Ortiz *et al.*, 2007).

This works proves that the different statistical techniques applied in combination with the phytosociological method can serve as an exploratory tool to foresee and suggest the best suitable formulae to adopt future decisions concerning the vegetal cover management for a sustainable agriculture.

Decision trees have been also used to classify and graphically represent the plant-soil relationships involved. A decision tree is a prediction model used in the field of artificial intelligence. Out of a given data base, logic diagrammes can be generated to represent and categorize a series of consecutively occurring conditions aiming at the solution of a problem.

A decision tree must have some inputs. These inputs may be either an object or a situation described by means of a series of features. After processing these inputs, the decision tree provides an output which can be used as a criterium to make a decision. The values of inputs and outputs can be either discrete or continuous. Discrete values are mostly used for simplicity reasons. When in the functions of an application discrete values are used, we call it *classification*; when it is continuous values that are used, we call it *regression*. Since we aimed at a classification, we used discrete values.

In this work the decision trees were generated by means of the data mining programme *Weka* (*http://www.cs.waikato.ac.nz/~ml/weka/*). For this purpose, we first created a series of data matrices with all the phytosociological relevés carried out, but including in each matrix only the Van der Maaler's indexes for a selected species together with all the values associated to edaphic features. Subsequently, the matrix was loaded in *Weka* and the information gain was then obtained as evaluator for the features. With this parameter we obtained a value indicating the information gain provided by each edaphic feature according to the occurrence of the relevant species. Subsequently we introduced the classifier (in our

Table I, Tableau I.- *Anacyclo clavati-Hordeetum leporinum*.

Order n°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of field rel.	124	125	126	127	128	129	130	131	132	133	134	135	137	138	139	140	141	142	143	136
Slope	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0
Exposure	-	-	-	-	-	-	-	-	-	-	S	S	-	-	-	-	-	-	-	-
Cover rate (%)	100	100	100	100	100	531	100	100	100	100	90	90	100	100	100	90	65	60	80	100
Area (m^2)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Altitude (m)	570	550	564	555	556	531	530	537	542	530	553	548	528	525	671	498	542	467	461	535
Veg. average height (m)	0.25	0.25	0.20	0.20	0.20	0.20	0.50	0.30	0.20	0.25	0.20	0.20	0.25	0.35	0.15	0.20	0.15	0.18	0.25	0.30
Dominant veg. average height	0.25	0.35	0.25	0.32	0.28	0.30	0.60	0.35	0.30	0.35	0.30	0.30	0.35	0.45	0.25	0.28	0.20	0.25	0.40	0.40
Characteristic species																				
Hordeum leporinum	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5	4	3	3	4	5
Sinapis alba subsp. *mairei*	1	1	1	+	+	+	+	+	+	+	+	+		+	+	+	1	+	1	
Erodium malacoides		+	+	+		1	+	+	+	1	+	1	1				+	1	+	1
Bromus diandrus	3	3	3	1	2	2	1	2	1	1	1	1	+	+						2
Anacyclus clavatus	2			+		+	+	+	1	+			+	1	+	1	1	+		1
Medicago polymorpha	+	+			+	+	1		+	+	+	+	1		+			+		1
Malva neglecta	+	+						1	1		+	+	2	1	1			+		2
Calendula arvensis		+	+	+				1	+		+		+			+		1	+	+
Filago lutescens	1		2	1	+	+	1	+	+				+							+
Bromus hordeaceus	2	1	1		+	+						1	+							+
Plantago lagopus			1				+		2	1			2			1				2
Leontodon longirrostris						+	+		+	+										1
Astragalus hamosus	1		+			+						1								+
Anagallis coerulea							+			+		+	1							+
Silene rubella								+	+				+					+		
Taraxacum erythrospermum										+				+						1
Picris comosa							+			+			+							
Medicago minima			1	+															+	
Geranium molle											1			+						1
Eruca sativa subsp. *longirostris*					+											1	1			
Erodium cicutarium			+			+														+
Erodium ciconium	+					+	+													
Convolvulus arvensis																1		+	+	
Bromus rubens											+	+					1			
Avena sterilis		+									+	+								
Veronica polita																		+	+	
Trifolium pratense						+	1													
Silene vulgaris																+	+			
Scorzonera laciniata																	+			+
Scandix pecten-veneris																		1	2	
Nonea vesicaria			+														+			
Lathyrus clymenum																		1	1	
Echium plantagineum							+		+											
Diplotaxis siifolia													1							2
Avena barbata										+				1						
Valerianella discoidea		+																		
Torilis nodosa							+													
Stachys arvensis																				+
Sonchus oleraceus																		+		
Sonchus asper												+								
Sisymbrium officinale							+													
Silene nocturna					+															
Senecio gallicus																		+		
Scandix australis		+																		
Plantago lanceolata																	+			
Papaver hybridum				+																
Ononis mollis				+																
Moricandia arvensis																	+			
Medicago coronata			+																	
Leontodon taraxacoides							1													
Hedypnois cretica																		+		
Galium aparine																		+		
Fumaria parviflora														+						
Euphorbia serrata		-																		
Euphorbia peplus															+					
Euphorbia helioscopia																		+		
Euphorbia exigua			+																	
Erodium botrys var. *brachycarpum*														+						
Diplotaxis catholica												+								
Centaurea pullata subsp. *baetica*																	+			
Capsella rubella																		+		
Bromus rigidus		+																		
Bromus racemosus		+																		
Avena fatua																	+			
Companion species																				
Carduus bourgaeanus				+	+		1		+		+				+					
Trifolium tomentosum			+					+												+
Lamarckia aurea											+	+	+							
Herniaria cinerea				1	1								1							
Foeniculum vulgare subsp. *piperitum*													+		+	1				
Conium maculatum										+	+	+								
Plantago albicans		+															+			
Cerastium glomeratum						+														-
Aristolochia paucinervis													+							-
Silybum marianum															1					
Scorpiurus muricatus		+																		
Oryzopsis miliacea														+						
Dactylis glomerata																				
Astragalus incanus subsp. *macrorhizus*																+				
Field n°	124	125	126	127	128	129	130	131	132	133	134	135	137	138	139	140	141	142	143	136
X-coordinate (m)	443529	443490	443491	443537	443521	442785	442773	443262	443246	443383	443672	443688	443833	442363	468736	472198	473134	476222	476174	443844
Y-coordinate (m)	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	418975	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06	4E+06

Area (m²)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Altitude (m)	188	264	355	373	355	365	281	278	279	278	267	491	491	618	620	664	625	302	295	196
Veg. average height (m)	0.20	0.30	0.35	0.15	0.30	0.20	0.35	0.22	0.25	0.30	0.30	0.30	0.30	0.20	0.25	0.25	0.30	0.35	0.25	0.20
Dominant veg. average height	0.32	0.42	0.43	0.25	0.45	0.25	0.50	0.35	0.40	0.45	0.40	0.45	0.400	0.25	0.35	0.40	0.40	0.50	0.35	0.30
Characteristic species																				
Hordeum leporinum	4	4	5	3	5	5	5	1	5	5	5	5	5	4	5	4	5	5	4	4
Geranium molle	3	+	1	2		3	1	1	+			+	1	1	1	2	2	1	+	
Bromus hordeaceus		+				1	2	+	+		1	2	+	+	+	+	+		+	2
Erodium moschatum	2	1	+		1	1	+	+						+	1		+	1		
Raphanus raphanistrum	+		+	+		+					+			+		+	+		1	
Medicago polymorpha	1	1	1	1	1	1												+	+	1
Plantago lagopus	1		+	1	1	+								1	1				1	
Echium plantagineum		+	+	1		+					+				+				+	1
Bromus diandrus	1	2	2				+			+	2					1	1			
Avena barbata							+	+	+	1	+	+						1	1	
Calendula arvensis	+	+		+						+			+						+	+
Anacyclus clavatus	+			1	+								+	+	+					+
Malva neglecta		1							+		1				+				+	
Bromus rubens		1					1	4									+			+
Leontodon longirostris											+	+				1				1
Bromus rigidus										+		2	1						+	
Sherardia arvensis		+	+										1							
Medicago arabica							+	+									+			
Euphorbia helioscopia		+	+															1		
Cynara humilis							1				+									+
Vulpia geniculata										1						1				
Stellaria media														+	+					
Ornithopus compressus														+	1					
Malva parviflora		1	+																	
Leontodon taraxacoides	+					+														
Hedypnois cretica			+					1												
Galium aparine									+									+		
Galactites tomentosa					+								1							
Capsella rubella														+					+	
Bromus scoparius														1	3					
Avena sterilis					3															+
Astragalus cymbicarpos	2	+																		
Anthriscus caucalis													+	+						
Vulpia membranacea																				+
Veronica arvensis																+				
Trifolium cherleri				+																
Sonchus oleraceus											+									
Sinapis alba subsp. *mairei*		1																		
Senecio gallicus																+				
Poa pratensis														1						
Nonea vesicaria		+																		
Lathyrus cicera														+						
Lathyrus angulatus				+																
Lagurus ovatus													+							
Geranium lucidum													+							
Galium parisiense					1															
Fumaria reuteri											+									
Fumaria parviflora																		+		
Erodium malacoides		+																		
Erodium cicutarium																				+
Erodium ciconium													+							
Diplotaxis virgata	+																			
Diplotaxis siifolia						+														
Crepis capillaris														+						
Convolvulus arvensis					2															
Chrysanthemum coronarium		+																		
Brassica barrelieri																	+			
Biserrula pelecinus															1					
Biscutella auriculata				+																
Anthemis praecox																				+
Anthemis arvensis																				1
Anagallis coerulea															+					
Companion species																				
Carduus bourgaeanus							+		+		+	1	+	1			+			
Foeniculum vulgare subsp. *piperitum*								+			+	+		+					+	
Trifolium subterraneum													1	1		1	1			
Paronychia argentea								+							+				+	
Plantago coronopus						+														1
Eryngium campestre								1		+										
Dactylis glomerata													+	+						
Trifolium tomentosum																				1
Silybum marianum	+																			
Poa infirma																+				
Phalaris minor																				
Hypochoeris radicata															+			+		
Geranium dissectum																		+		
Equisetum ramosissimum																		+		
Cynodon dactylon								1												
Cerastium pentandrum														2						
Cerastium glomeratum																+				
Cerastium brachypetalum																	+			
Carduus pycnocephalus	1																			
Field n°.	118	119	120	121	122	123	144	145	146	147	148	149	150	151	152	155	156	157	158	161
X-coordinate (m)	394356	393051	388213	388428	388827	388483	412499	412481	412542	412655	412678	414904	415436	413745	413614	409242	409094	417850	417652	395366
Y-coordinate (m)	4212231	5210941	4217611	4217221	4216674	4217146	4213612	4213583	4213608	4213889	4213928	4217234	4219003	4219563	4219466	4219500	4219081	4217820	4217488	4215177

he algorithm J48) to generate the decision tree, but including only the relevant eda-·atures, that is those whose information gain was not null and, consequently, not only ed relevant information but also fail to produce noise.

is, each of the decision trees so generated performs a test when moving through the ies to finally come to a decision. The represented decision trees contain inner nodes, odes and arcs. An inner node contains a test of a value of one of the properties (in our he range of values of an edaphic feature). A leaf node represents a value returned by ecision tree (the Van der Maaler index for that particular species). Finally, the ies provide all possible ways available according to the decision taken.

III. RESULTS AND CONCLUSIONS

ytosociological analysis

clo clavati-Hordeetum leporinum ass. nova *(AH)*

he territories of the Guadalquivir river valley (Hispalensean Sector) and Subbetic there is a *Hordeum leporinum* pastureland which develops on basic substrates with value higher than 8. These pasturelands frequently include basophilous species such *dicago minima, Sinapis alba* subsp. *mairei, Moricandia arvensis, Centaurea pulla-ntago albicans*, which are absent in the *Bromo scoparii-Hordeetum leporinum* asso-, described for the continentalised areas of the Meseta (Rivas-Martínez, 1978). The ncy of *Anacyclus clavatus*, a more or less edaphically indifferent taxon, but with a ence for basic substrates, and the absence of *Anacyclus radiatus*, a thermophilous nt which does not reach to the Alto Guadalquivir, lead us to propose the new asso- *Anacyclo clavati-Hordeetum leporinum nova* (Table I, rel. 1 al 20; *typus* rel. 9). In tistical analysis this association appears perfectly differentiated both from *Bromo rii-Hordeetum leporinum* and *Anacyclo radiati-Hordeetum leporinum* (Fig. 1).

distribution area of this syntaxon is located in the mesomediterranean, dry-subhu-lt of the Betic territories, in the area of the *Paeonio-Querco rotundifoliae* s., *Viburno uerco fagineae* s. series.

scoparii-Hordeetum leporinum *Rivas-Martínez 1978 (BH)*

s pastureland, whose optimum stage is located in the meso- and supramediterranean otypes, presents a therophytic character and vernal phenology. It was described for ntinentalised environments of the centre of Spain (Rivas-Martínez, 1978; Cano-Ortiz 2009). The association has the following catenal contacts: towards scarcely nitrified y soils, with the *Taeniathero-Aegilopion geniculatae* pasturelands; when the mois-vel is higher, with the vallicares of *Agrostion salmanticae*; finally, towards more d areas, with the *Malvenion neglectae* grasslands.

tistical analysis

results of the cluster analysis revealed that the general multivariate analysis of the lots under study allows to suitably group these land plots according to their phyto-ogical features (Fig. 1). Only in the land plots of the pastureland BH association and, e cases, in the land plots of the herbaceous AH association, as a result of the pre-of basophilous elements such as *Moricandia arvensis, Medicago minima, Plantago ns, Anacyclus clavatus*, can be found little grouping abnormalities. On the other the two-block or tandem distribution of the associations *Papaveri rhoeadis-*

Diplotaxietum virgatae (PD) and *Urtico urentis-Malvetum neglectae* (UM) is due to the presence of *Malva parviflora* and *Hordeum leporinum* in the land plots closest to AH and BH.

The study of the relationships between the different associations sampled according to their floristic composition was also carried out by means of graphical networks of co-location. The data matrices only included, for each association, a group of characteristic species selected according to their presence in a given number of samples, their abundance-dominance value and ecology. The relationship between associations phytosociologically close according to the shared species was also studied.

In Table III we indicate with an X the presence of a certain species in the corresponding associations. These are (AH) *Anacyclo clavati-Hordeetum leporinum* and (BH) *Bromo scoparii-Hordeetum leporinum*. Some species can be found in both associations, but the rest can only be found in one of them. This is no surprise because some species have peculiar edaphic requirements and must, consequently, react to less wide ranges than those imposed by the edaphic parameters under study. A comparative analysis of the associations (AH-BH) reveals the presence of shared and different taxa. The checking, by means of a correlation network, of the association *Anacyclo clavati-Hordeetum leporinum* (AH) against the *Bromo scoparii-Hordeetum leporinum* association (BH) reveals that there is a group of plants shared by both associations (*Medicago polymorpha, Bromus diandrus, Bromus hordeaceus, Hordeum leporinum*). Meanwhile, the whole number of species shared by both associations appear in two groups (Fig. 2) due to a remarkable difference as far as the edaphic parameters, such as P, K, pH, etc. are concerned. As the statistical tables obtained with robust methods (Fig. 3) clearly show, these factors may determine the existence of these associations.

Table III.- Presence of species.
Tableau III.- Présence des espèces.

	AH	BH
Medicago polymorpha	X	X
Malva neglecta	X	.
Calendula arvensis	X	.
Erodium malacoides	X	.
Bromus diandrus	X	X
Hordeum leporinum	X	X
Sinapis alba subsp. *mairei*	X	.
Erodium moschatum	.	X
Bromus hordeaceus	X	X

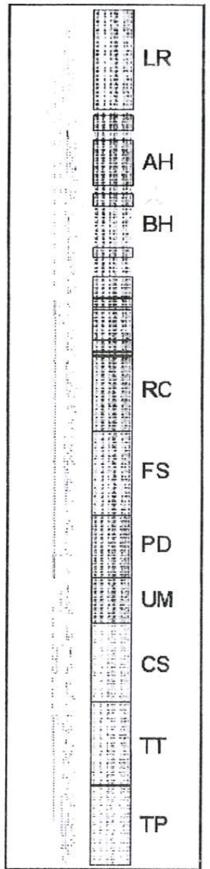
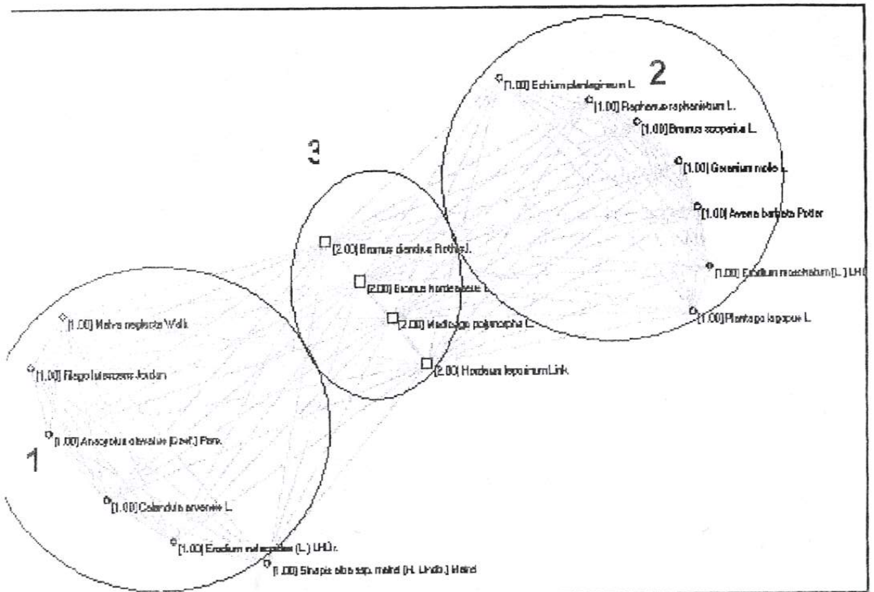


Fig. 1.- Statistical analysis (cluster).
Fig. 1.- Analyse statistique (cluster).

After the multiple phytosociological and edaphic analyses carried out for the communities of *Hordeum leporinum* (*Bromo scoparii-Hordeetum leporinum* (BH), *Anacyclo clavati-Hordeetum leporinum* (AH)) and the multiple statistical techniques used to determine the correlation between these data, let us now comment on the descriptive statistical arrangement for these communities.

In Tables VI and VII, we show the descriptive statistical arrangement obtained for the communities under study. These are: mean, median, variance, standard deviation, standard error, minimum, maximum, range, first quartile, second quartile and coefficient of variation. The arrangements reveals that the community of *Hordeum leporinum* AH presents a high CEC value, with a median of 15.8245 meq/100 g. The pH-value is higher than 8 and the OMM-value is 1.455%. The texture is sandy-silty and, consequently, the retentive capacity is moderate. The community of *Hordeum leporinum* BH presents a pH-value of



2.- Phytosociological co-location network of the associations. 1 - Species of *Anacyclo avati-Hordeetum leporinum* (AH). 2. Species of *Bromo scoparii-Hordeetum leporinum* IH). 3 - Species shared by both associations (Cano-Ortiz, 2007).

2.- Réseau d'emplacement phytosociologique des associations. 1 - Espèces de *Anacyclo clavati-Hordeetum leporinum* (AH). 2 - Espèces du *Bromo scoparii-Hordeetum porinum* (BH). 3 - Espèces communes aux deux associations (Cano-Ortiz, 2007).

5 and an OMM-value of 1.475%. The CEC-value, 9.347 meq/100 g, is in this case iarkably lower than that of the former community. The texture is sandy-silty and the :ntive capacity is moderate.

The high heterocedasticity revealed by the data (as a result of extreme atypical values) l shown in the tables (note the differences between means and medians) led us to implent statistically robust methods to analyse and compare the edaphic features (Cano-Ortiz,)7).

Since the classical analysis reveals not only the presence of anomalous data but also a h heterocedasticity for the associations involved, statistically robust methods must be plied to describe and compare those edaphic features according to the association corponding to the relevés in question.

As robust estimator for the mean we used the sample Winsorized mean (media-win), ich avoids the high sensitivity to the presence of extreme values. As robust estimator for : median we used the Huber M-estimator of location (mediana-est), an estimator which ll be coincidental with the sample mean if the data of the model are not anomalous, since : sample mean is actually the most reliable estimator. As shown in the tables and figure the mean and the mediana-est for the pH-values are almost coincidental, which indicates it there are no anomalous data. However, this is not the case as far the rest of parames are concerned, which indicates that there are anomalous data (Fig. 3).

Table IV.- Edaphic table of *Anacyclo clavati-Hordeetum leporinum*.
Tableau IV.- Tableau édaphique de l'*Anacyclo clavati-Hordeetum leporinum*.

NUMRILCAMP	124	125	126	127	128	129	130	131	132	133	134	135	137	138	139	140
ID Rilievo	276	277	278	280	282	283	284	285	286	287	288	289	290	291	292	293
C.I.C (meq/100g)	16.087	13.043	13.043	20	11.304	15.562	17.391	15.217	16.522	13.478	15.957	16.087	18.696	11.304	13.913	[illegible]
Carbonatos (%)	52.5	55.9	56.2	51.6	46.2	45.9	40.5	28.8	29.6	29.6	38.6	36.7	42.5	52.5	67.3	[illegible]
Ca (meq/100g)																
P asimilable (p.p.m)	7	5	6	7	5	7	25	9	11	8	9	9	11	13	27	[illegible]
Mg (meq/100g)	1.326	1.176	0.898	0.967	1.026	1.941	2.373	2.274	1.699	1.442	1.495	1.732	1.735	1.041	2.175	[illegible]
M.O.O (%)	1.76	1.47	1.42	1.41	1.36	1.13	2.72	0.98	0.82	1.15	1.53	1.58	1.44	2.23	2.25	[illegible]
N (%)	0.122	0.105	0.092	0.104	0.105	0.089	0.194	0.075	0.066	0.095	0.118	0.107	0.126	0.158	0.155	[illegible]
pH 1/2,5	8.4	8.3	8.4	8.4	8.2	8.5	7.9	8.3	8.1	8.4	8.4	8.4	8.4	7.6	8.2	[illegible]
K (meq/100g)	0.563	0.46	0.512	0.563	0.409	0.614	0.87	1.483	0.409	0.46	0.87	0.818	1.586	0.409	0.818	[illegible]
pF 1/3 atm (%)	29.86	30.32	26.36	30.35	28.77	16.14	27.53	26.1	28.79	23.91	26.66	26.32	31.74	19.6	32.39	[illegible]
pF 15 atm (%)	15.79	11.92	13.74	16.57	15.95	14.56	16.64	15.18	15.18	11.85	14.1	14.76	19.15	11.18	17.92	[illegible]
Tx arcilla (%)	16.55	27.2	9.1	5.1	6.15	6.05	8.55	35.9	30.75	8.3	27.45	4.5	3.65	21.45	30.55	[illegible]
Tx arena (%)	10.2	28.11	13.8	11.74	11.68	13.34	18.04	21.97	16.52	33.26	24.83	23.33	13.89	42.92	25.45	[illegible]
Tx limo (%)	73.25	44.69	77.1	83.16	82.17	80.61	73.41	42.13	52.73	58.44	47.72	72.17	82.46	35.63	44	[illegible]
Tamiz 2mm (%)	34.8	32.1	34.09	19.81	24.84	46.14	55.65	56.87	32.16	34.7	38.83	37.63	47.67	31.84	20.4	[illegible]
Salinidad (mmhos/cm)	0.2	0.2	0.2	0.19	0.3	0.19	0.56	0.5	0.42	0.19	0.18	0.16	0.29	2	0.22	[illegible]

Table V.- Edaphic table of *Bromo scoparii-Hordeetum leporinum*.
Tableau V.- Tableau édaphique du *Bromo scoparii-Hordeetum leporinum*.

NUMRILCAMP	118	119	120	121	122	123	144	145	146	147	148	149	150	151	152	155
ID Rilievo	279	281	297	298	299	300	301	302	303	304	305	306	307	308	309	310
C.I.C (meq/100g)	7.826	15.217	13.913	10.13	13.913	21.739	9.565	10	11.304	8.697	6.957	8.697	9.13	6.087	5.652	9.565
Carbonatos (%)	10.2	34.2	2.2	2.4	6.5	3.4	3.3	2.1	3.6	2.3	2.5	2.5	1.1	1.5	1.6	2.5
Ca (meq/100g)			22.128	34.691	35.1	54.491	22.567	15.783	33.123	16.726	15.79	15.43	17.5	9.59	10.988	12.702
P asimilable (p.p.m)	4	49	4	3	4	14	7	1	109	1	1	0	3	1	12	[illegible]
Mg (meq/100g)	0.439	0.71	0.401	3.104	0.681	1.123	0.735	0.89	1.589	1.188	0.882	0.704	0.602	0.686	0.467	[illegible]
M.O.O (%)	0.59	1.68	1.75	0.81	1.92	7.21	1.78	1.38	2.85	1.22	1.03	2.16	1.57	0.74	1.59	[illegible]
N (%)	0.052	0.111	0.143	0.1	0.17	0.455	0.148	0.128	0.274	0.112	0.094	0.127	0.109	0.049	0.115	[illegible]
pH 1/2,5	8	8.2	7.8	7.9	8	7.4	8	7.4	7.8	7.7	7.8	7.3	7.6	6.4	6.9	[illegible]
K (meq/100g)	0.512	1.739	0.095	0.072	0.128	0.133	0.409	0.136	1.535	0.11	0.358	0.24	0.248	0.217	0.358	[illegible]
pF 1/3 atm (%)	15.03	27.4	29.04	28.07	28.55	40.94	22.95	18.98	22.4	19.7	15.31	19.06	10.44	12.77	13.75	[illegible]
pF 15 atm (%)	6.03	11.46	11.59	15.2	10.26	17.79	10.19	7.79	2.25	6.59	5.41	6.76	7.47	5.71	[illegible]	7.54
Tx arcilla (%)	17.25	18.85	21.2	27.65	14.85	19.65	20.05	14.3	17.35	11.45	9.25	11.15	11.5	8.95	[illegible]	[illegible]
Tx arena (%)	65.29	46.44	35.19	27.17	39.88	26.18	42.51	54.34	53.35	50.96	68.81	55.94	64.81	73.64	[illegible]	[illegible]
Tx limo (%)	17.46	34.71	43.61	45.18	45.28	54.17	36.84	31.36	29.3	37.59	21.94	32.91	23.69	17.41	[illegible]	[illegible]
Tamiz 2mm (%)	5.97	9.11	24.39	41.06	22.59	15.88	28.8	23.99	24.09	6.93	9.4	17.15	7.74	[illegible]	[illegible]	[illegible]
Salinidad (mmhos/cm)	0.14	0.21	0.12	0.15	0.15	0.24	0.15	0.06	0.19	0.09	0.15	0.09	0.09	0.05	[illegible]	[illegible]

As already commented in the section *Materials and methods*, the quantitative approach of the Bayesian statistics relies on the fact that each variable of the model has a limited number of possible states (or levels) and each one is associated to a value, the probability of that state.

When we have some previous evidence about any of the variables of the model, the probabilities can be accordingly modified and, in that case, when we know the value which a given variable of the model has, the information propagates through the network, and the

Mg (meq/100g)	AH	BH
media	1.683	1.068
media-win	1.558	0.844
mediana	1.595	0.723
mediana-est	1.570	0.818
nqse	0.204	0.110
mestse	0.133	0.069

M.O.O. (%)	AH	BH
media	1.54	1.67
media-win	1.46	1.36
mediana	1.46	1.48
mediana-est	1.48	1.40
nqse	0.15	0.21
mestse	0.11	0.15

N (%)	AH	BH
media	0.115	0.138
media-win	0.110	0.125
mediana	0.105	0.112
mediana-est	0.110	0.115
nqse	0.008	0.012
mestse	0.008	0.009

pH	AH	BH
media	8.3	7.5
media-win	8.3	7.6
mediana	8.4	7.8
mediana-est	8.3	7.6
nqse	0.0	0.2
mestse	0.0	0.2

pF 15 atm (%)	AH	BH
media	15.32	8.20
media-win	15.33	7.69
mediana	15.49	7.51
mediana-est	15.42	7.75
nqse	0.70	0.88
mestse	0.65	0.81

Tx arcilla (%)	AH	BH
media	17.76	14.50
media-win	16.49	14.15
mediana	12.83	13.33
mediana-est	16.65	14.21
nqse	5.07	1.61
mestse	3.59	1.12

Tx arena (%)	AH	BH
media	20.45	54.25
media-win	18.51	55.18
mediana	18.25	53.85
mediana-est	19.37	54.70
nqse	2.41	5.60
mestse	2.24	3.75

Tx limo (%)	AH	BH
media	61.79	31.24
media-win	61.99	30.85
mediana	59.62	32.14
mediana-est	61.79	31.08
nqse	6.90	4.71
mestse	3.83	2.71

Tamiz 2mm (%)	AH	BH
media	35.41	19.04
media-win	36.92	18.03
mediana	34.75	19.03
mediana-est	35.95	18.61
nqse	4.35	2.89
mestse	2.68	2.30

Salinidad (mmhos/cm)	AH
media	0.36
media-win	0.25
mediana	0.22
mediana-est	0.25
nqse	0.03
mestse	0.03

Fig. 3.- Application of statistically robust methods to several edaphic parameters (AH, BH).
Fig. 3.- Application de statistiques robustes à différents paramètres édaphiques (AH, BH).

values associated to other variables are recalculated (in the so-called *evidence propagation*).

In the graphical representation of points of probability on a Bayesian network the nodes stand for variables and, by means of statistically comparative processes, direct dependances between variables can be found. A table of conditioned probabilities is thus generated to describe the probability of the state of each node. As a result, the multiple combinations of states of the parentals nodes involved indicate how intense the causal relationships between variables are.

In our case, we have generated the directed acyclic graphs and the tables of probabilities for the species conferring the name to each association and for those species providing the largest amount of information in each of them. When selecting the highest Van der Maaler index for each species, we took into consideration that in over 80% of the cases the values of the edaphic parameters involved are within well-defined ranges.

Tables VI, VII.- Descriptive statistical arrangement for the BH and AH.
Tables VI,VII.- Statistiques descriptives pour les associations BH et AH.

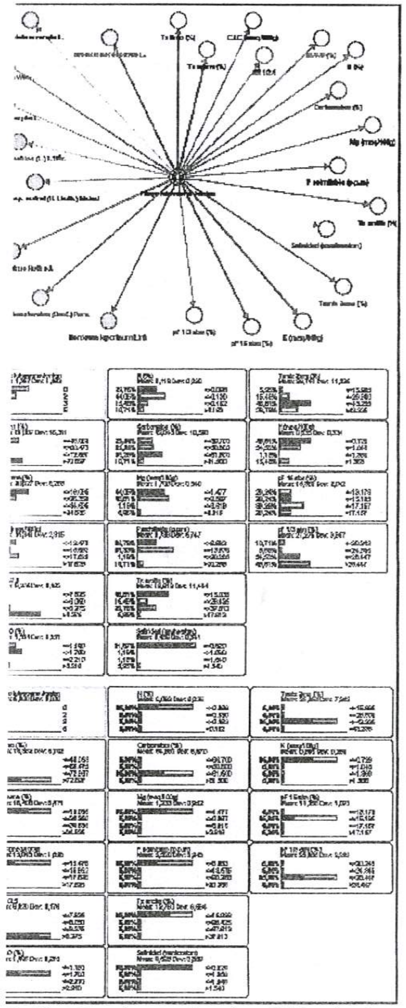
[illegible]	[illegible]

[illegible]	[illegible]

In the case of the *Anacyclo clavati-Hordeetum leporinum* association, the Bayesian networks corresponding to the species *Hordeum leporinum, Filago lutescens* are shown. In the case of *Filago lutecens*, before any evidence propagation, it is clear that, generally speaking, most of the edaphic parameters present a wide range and that this species can also occur with a different Van der Maaler index. However, when the previous evidence that the species occurs with the highest abundance rate (9 in this case) is established, the evidence propagates and, as a result, we obtained a characteristic and bounded distribution for many of the soil features, which allows us to detect its necessity in values determined for that species (Fig. 4).

The species *Hordeum leporinum* is initially found in 68.25% of the cases with a Van der Maaler index of 9, in 20.63% with an index of 8 and in 11.11%, of 7. With these abundance values for this species, the edaphic parameters fluctuate between some concrete ranges. For some edaphic parameters the following data were recorded: in 91.67% of the cases salinity shows values lower than or equal to 0.620 mmhos/cm; in 44.05% of the cases H shows values lower than or equal to 0.130%; in 58.33% of the cases P shows values lower than or equal to 13.575 p.p.m; and in 48.81% of the cases K show values lower than or equal to 0.729 meq/100 g. When we establish the previous evidence that the species has a Van der Maaler index of 9, among the previously mentioned edaphic parameters it is mainly K that most heavily varies in proportion: in 63.37% of the cases it even shows a value of 0.729 meq/100 g (Fig. 5).

Among the procedures of expert systems used, one of the classification systems applied to find out plant-soil relationships has been the generation of decision trees for the most representative species, *Hordeum leporinum*. We first measured the information gain as evaluator of features, and subsequently introduced the algorithm of classification (J48) to generate the tree. Since in some cases the values recorded for the information gain of all



4, Fig. 4.- *Anacyclo clavati-Hordeetum* ›orinum: *Filago lutescens*.

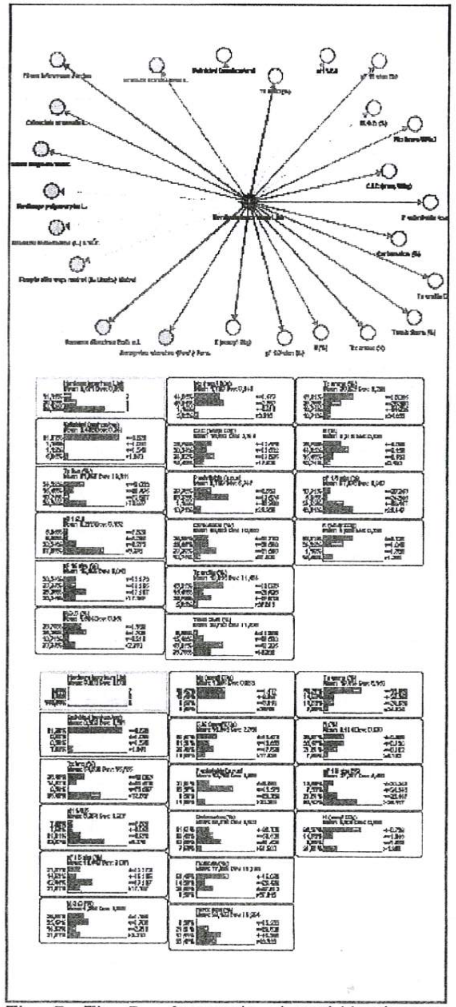


Fig. 5, Fig. 5.- *Anacyclo clavati-Hordeetum leporinum*: *Hordeum leporinum*.

›eatures of a certain characteristic species were equal to null, we show here only those ›ifiers whose information gain for the features being represented is different from null. ›formation gain (IG) can be defined as the amount of information gained when selec- a feature (parameter edaphic) and using it to classify a class (plant species) according Consequently, the higher the IG value, the more important the edaphic parameter is ›und the abundance of the species, and the higher the number of edaphic parameters IG, the more demanding the species is when colonizing this or that soil.

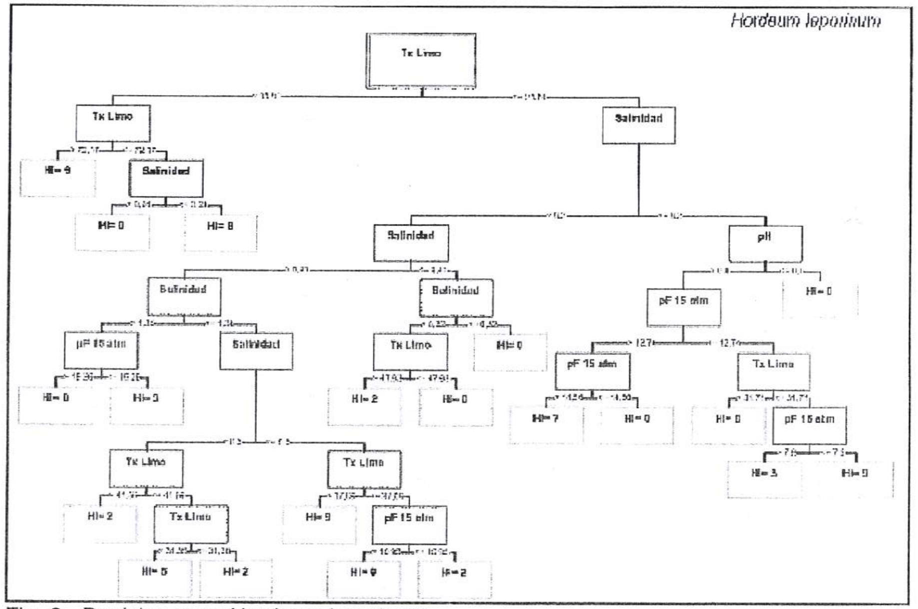


Fig. 6.- Decision tree: *Hordeum leporinum*.
Fig. 6.- Arbre de décision : *Hordeum leporinum*.

Below, we can see, for the species *Hordeum leporinum*, the value of information gain per edaphic feature and the classifier decision tree for the abundance-dominance index of the species according to those features.

Hordeum leporinum.
Features ranked by information gain:

0.285	15 salinity
0.148	13 Tx_silt
0.119	7 pH1
0.11	10 pF_15atm
0	5 OMM
0	4 Mg
0	1 CEC
0	3 P_assimilable
0	2 carbonates
0	11 Tx_limestone
0	14 Sieve_2mm
0	12 Tx_sand
0	6 N
0	9 pF_1_3atm
0	8 K

hic features having the greatest impact on the classification of this species are Tx silt, pH and pF 15 atm, salinity being the most influential of them all. In the tree corresponding to this species we can see that we can get to a Van der Maaler ce index of 9 for *Hordeum leporinum* through different ways: from a Tx silt higher 7% we get to an abundance rate of 9 for this species 9; meanwhile, if the per- of Tx silt is higher than 53.81 and lower than or equal to 72.17 and salinity is an 0.24 mmhos/cm, the species is missing (Fig. 6).

REFERENCES

juet J., 1979.- *Fitosociología. Bases para el o las comunidades vegetales.* Ed. H. Blume,

A., 2007.- *Bioindicadores ecológicos y mane- biertas vegetales como herramienta para la ción de una agricultura sostenible.* Tesis Universidad de Jaén.

A., F.J. Esteban Ruiz, C.J. Pinto Gomes, A. z Torres, J. Goñi, I. De la Haza & E. Cano, aracterización de comunidades agrofores- ílisis multivariable y bayesiano. *Cuad. Soc. ic. For.*, **22**, 25-29.

A., C.J. Pinto-Gomes, F.J. Esteban Ruiz, A. z-Torres, J. Goñi, I. De la Haza & E. Cano, odiversity of *Hordeion leporini* in Portugal: a iological and edaphic analysis. *Acta Bot.* **56** (1), 33-49.

S. *et al.* (eds), 2001.- *Claves de Flora* CSIC, 774 p.

García Pérez A., 2005.- *Métodos avanzados de estadística aplicada: métodos robustos y de muestreo.* UNED, 255 p.

Knoepp J.D. *et al.*, 2000.- Biological indices of soil quality: an ecosystem case study of their use. *Forest Ecol. Manag.*, **138**, 357-368.

Pereira Coutinho A.X., 1939.- *Flora de Portugal (plantas vasculares).* Ed. Bertrand, Lisboa, 938 p.

Rivas-Martínez S., 1978.- La vegetación de *Hordeion leporini* en España. *Doc. Phytosociol.*, **XI**, 377-392.

Rivas-Martínez S., 1996.- Clasificación bioclimática de a Tierra. *Folia Bot. Matritensis*, **16**, 1-32.

Tutin T., V.H. Heywood, D.A. Burges, D.H. Valentine, S.M. Walters & D.A. Webb (eds), 1964-93.- *Flora Europea*, I-IV, Cambrige University Press.

Valdés B., S. Talavera & E. Fernández-Galiano, 1987.- *Flora Vascular de Andalucía Occidental.* Ketres, **1-3**, Barcelona.

Acta Bot. Gallica, 156 (4), 625-636, 2009.

Recent climatic trends in Castilla and León (Spain) and its possible influence on the potential natural vegetation

by Sara del Río, Luis Herrèro and Àngel Penas

Departamento de Biodiversidad y Gestión Ambiental (Área de Botánica), Facultad de Ciencias biológicas y ambientales, Universidad de León, Campus de Vegazana s/n., E-24071 León; sriog@unileón.es

Abstract.- Precipitation and mean temperature trends at monthly, seasonal and annual time scale for Castilla and León during the period 1961-2004 were analysed in the present study. The main aims of this paper were to find out if these tendencies differ from those pointed out for this region in a previous research for the period 1961-1997 and to establish, based on bioclimatic models, possible changes on the potential natural vegetation of this territory if these trends do not change in future. Magnitude of trends was derived from the slopes of the regression lines while the statistical significance was determined by the Mann-Kendall test. Magnitude and sign of rainfall and temperature trends over the last four decades have been essentially the same than those concerning to the period 1961-1997. It can be concluded that if these trends keep in future areas with temperate climate could increase in this region. In this situation, deciduous forests could extend their distribution limits, thus replacing some semi-deciduous and evergreen ones.

Key words : bioclimatology - Castilla and León - climatic trends - potential natural vegetation - Spain.

Résumé.- On examine ici les tendances de précipitation et température moyenne mensuelle, saisonnière et annuelle pour la plus grande région de l'Espagne (Castilla et León) pendant la période 1961-2004. Le première objectif était découvrir si ces tendances ont été différentes de celles précisées pour cette région dans une recherche précédente pour la période 1961-1997 et en même temps d'établir, en fonction des modèles bioclimatiques, les changements possibles sur la végétation potentielle de ce territoire. L'intensité de cette tendance a été déduite des pentes de lignes de régression par l'utilisation de la méthode des moindres carrés et la signification statistique a été aussi déterminée par le test non paramétrique Mann-Kendall. Le sens des tendances de précipitations au cours des quarante-quatre dernières années est le même que celui de la période 1961-1997. En conclusion, si ces tendances continuent dans le futur, les secteurs de climat tempéré pourraient augmenter dans cette région. En conséquence, les forêts à feuilles caduques pourront accroître leur aire en remplaçant un certain nombre de forêts à feuilles semi-caduques et persistantes.

Mots clés : bioclimatologie - Castilla et León - tendances climatiques - Espagne - végétation potentielle.