ANALYSIS OF DIFFERENT STATISTICAL MODELS FOR ASSESSING POTENTIAL DISTRIBUTION OF FOREST TYPES IN SOUTHERN SPAIN

M. Anaya Romero[1,2]; R. Pino[3]; A. Jordán[2]; N. Bellinfante[2]; L. Martínez-Zavala[2]; & I. Gómez[2]

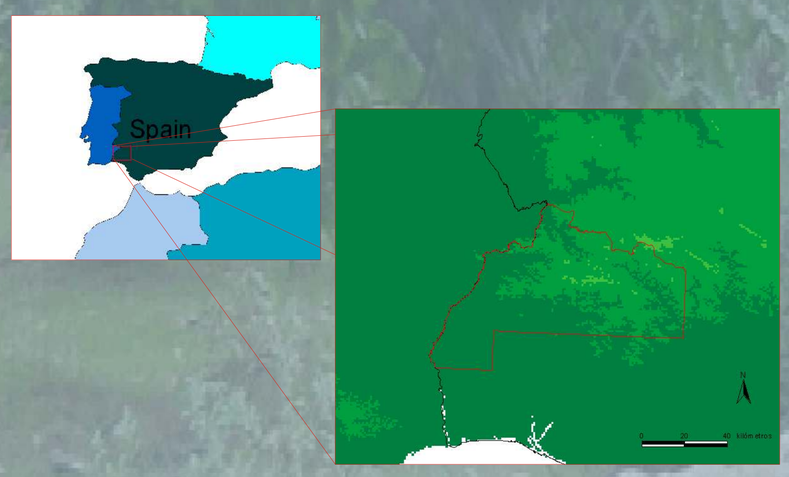
[1]Consejo Superior de Investigaciones Científicas (CSIC), Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), Avda. Reina Mercedes 10, 41012 Sevilla (Spain)
[2]Departamento de Cristalografía, Mineralogía y Química Agrícola (Universidad de Sevilla). C/ Profesor García González 1. C.P.: 41012 Sevilla (Spain)
[3]Departamento de Estadística e Investigación Operativa (Universidad de Sevilla). C/ Tarfia s/n. C.P.: 41012 Sevilla (Spain)

anaya@irnase.csic.es

18th World Congress
of Soil Science
July 9-15, 2006.
Philadelphia, Pennsylvania
USA

INTRODUCTION

The accomplishment of models to evaluate forest systems using environmental information has been used by different authors in very diverse ways. Nevertheless, the accuracy of the models is very influenced by the quality and number of variables used, as by the mathematical and statistical base used. Following Anaya Romero (2004), the election of the more accurate prediction model is approached in the present work, applied to forest system evaluation. The main goal of this evaluation is to predict the potential distribution of several forest types in Sierra de Aracena Natural Park and Western Andévalo (Huelva, Spain). The selected models relate the presence/absence of each forest type to the main parameters influencing the distribution of forest species in the Mediterranean environmental conditions of the study area. The models were selected taking into account their great predictive and explanatory capacity: Logistic Regression (LR), Artificial Neuronal Network (ANN) and Decision Tree (DT), Random Forest, Support Vector Machines (SVM).



METHODS

The forest types in the area are oak forest (Quercus suber and Q. rotundifolia), pine tree forest (Pinus pinaster and P. pinea), eucalyptus forest (Eucalyptus globulus and E. Camaldulensis) and deciduous forest. The selected environmental variables were grouped in several thematic categories: litology (type of rock, acidity and consolidation), geomorphology (erosive processes, mass movements, sedimentation processes and morphogenesis), physiography, relief (elevation, slope and hillslope facing), soil (pH, soil nutrients, organic matter, CEC and clay content) and climate (average summer precipitation, annual average temperature, average temperature of the warmest month and average temperature of the coldest month). After the exploratory analysis of the data, a sampling was made on which the selected models of prediction were applied. With the objective of studying the behaviour of each prediction analysis when new data were used, the whole information was divided in training-data and test-data. The division was made randomly, so that 75% of the samples were classified as training-data, and 25% as test-data. The results obtained by the five prediction models were compared using the Kleinbaum confusion matrix. Finally, DT was the method with a lowest Error Index (0.08 %).

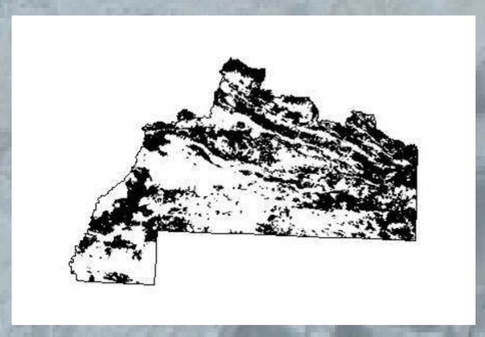
Oak forest

Pinus forest

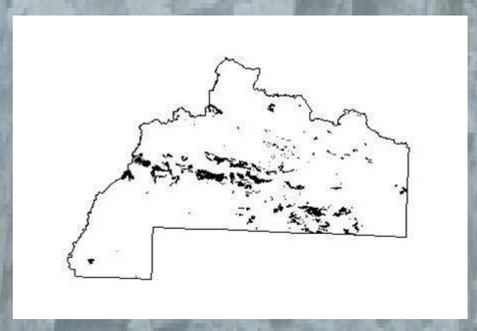
Broad leaved forest

Eucalyptus forest

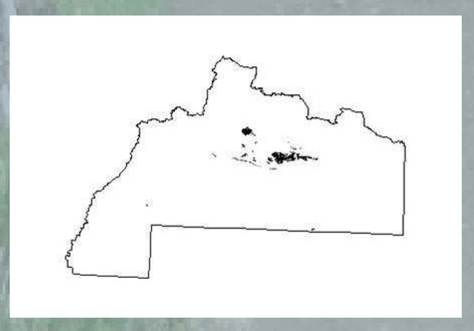
EXPLORATORY ANALYSIS OF DATA



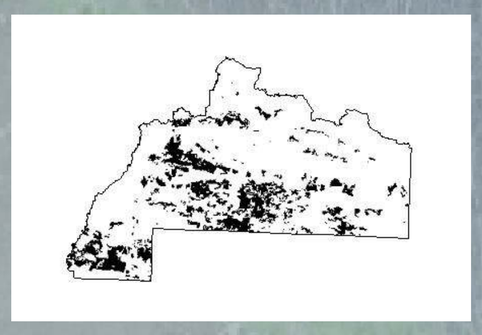
Oak forest



Pinus forest



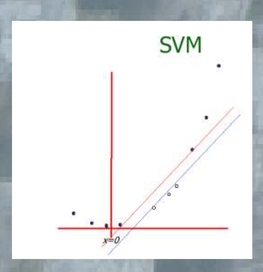
Broad leaved forest



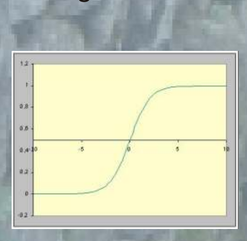
Eucalyptus forest

APPLICATION OF EVALUATION MODELS

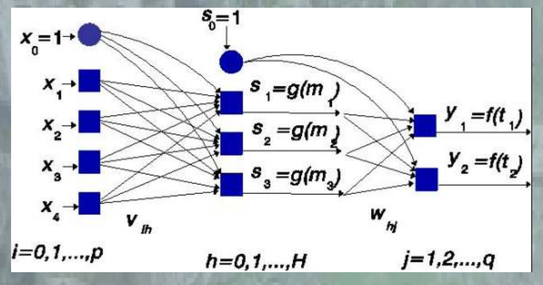
Support Vector Machine



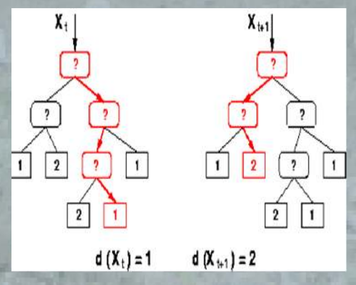
Logistic Regression



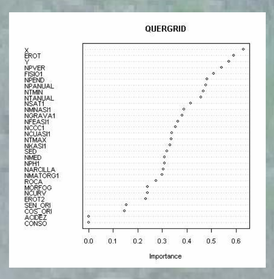
Artificial Neural Network



Decision Tree



Random Forest



VALIDATION OF MODELS

The applied models were validated using the Kleinbaum confussion matrix:

Index of estimation error (Kleinbaum confusion matrix)

		Predicted values		
		No	Yes	Total
Real Values	No	F11	F12	T1
	Yes	F21	F22	T2
	Total	T3	T4	T

Error index = (F12 + F21/ T) x 100

Index of estimation success: [(F21+F12)/T]x100

	Support Vector Machines	Logistic regression	Artificial neuronal network	Decision tree	Random Forest
Quercus	78,14	75,24	75,86	81,10	81,12
Pinus	96,19	96,43	3,78	96,29	89,25
Broad Leaved Formation	99,13	98,81	0,92	98,78	98,59
Eucaliptus	85,86	84,52	84,99	88,06	81,21