COMPARISON OF CLASSIFICATION METHODSUSED FOR PREDICTING GRASSHOPPER POPULATION

ISAIAS CHAIREZ HERNANDEZ

ADVISOR: TONGHUI WANG

Department of Mathematical Sciences

New Mexico State University

Abstract: Grasshopper population per square meter was assessed in grassland area of Durango México. Since grasshopper per square meter did not hold normal distribution, generalized linear models were used to interpolate data where grasshopper population per square meter was the dependent variable and ecological variables as independent variables. In this paper, Several methods such as the performance of multiple nonparametric discriminant analysis, kernel discriminant analysis (KDA), logistic regression (LR), and artificial neural networks (ANN) were used to compare in predicting quartiles categories (QC) and five size categories (FSC) of grasshopper per square meter (Training set 50% and testing set 50%). Results showed that the generalized linear model that best fitted survey data was Negative binomial with the natural logarithm as a link function and R^2=0.91. Uniform, Normal, and Epanechnikov kernels with radium from 0.001 to 1.0 were performed with KDA. Results established that Normal kernel was the most appropriate. ANN model, had the best performance in the FSC category, followed by KDA and LR. Grasshopper is part of the food chain in grassland ecosystems but in special ecological conditions the population increases and may cause damage to grassland and crops. Both ANN and KDA are efficient methods for population threshold predictions. Variable importance analysis showed that longitude, soil and precipitation were the variables that most influenced grasshopper per square meter in KDA and ANN analysis. Vegetation was important only in ANN.