Geophysical Research Abstracts
Vol. 19, EGU2017-12081-3, 2017
EGU General Assembly 2017


# Applying data mining methods to the assessment of soil contamination and carbon sequestration under Mediterranean Climate. The case study of Guadiamar basin (SW Spain).

Sara Muñoz Vallés (1), Rafael Pino-Mejías (2), Francisco J. Blanco-Velázquez (1), and María Anaya-Romero (1)

(1) Evenor-Tech "Technology-Based Company focus on Solutions for Soil Use and protection". Sevilla, Spain (info@evenor-tech.com), (2) Department of Statistics and Operational Research, University of Seville, Avda. Reina Mercedes s/n, Seville, Spain.

In the present background of increasing access to vast datasets of soil and environmental records, the application of the newest analytical techniques and approaches for modelling offer excellent opportunities to define recommendations and simulate processes for land degradation and management. In this regard, data mining techniques have been successfully applied in different fields of environmental sciences, performing an innovative tool to explore relevant questions and providing valuable results and useful applications through an efficient management and analysis of large and heterogeneous datasets.
Soil Organic matter, pH and trace elements in soil perform close relationships, with ability to alter each other and lead to emerging, synergic properties for soils. In addition, effects associated to climate and land use change promotes mechanisms of feedback that could amplify the negative effects of soil contamination on human health, biodiversity conservation and soil ecosystem services maintenance. The aim of this study was to build and compare several data mining models for the prediction of potential and interrelated functions of soil contamination and carbon sequestration by soils. In this context, under the framework of the EU RECARE project (Preventing and Remediating degradation of Soils in Europe through Land Care), the Guadiamar valley (SW Spain) is used as case study. The area was affected by around four hm3 of acid waters and two hm3 of mud rich in heavy metals, resulting from a mine spill, in 1998, where more than 4,600 ha of agricultural and pasture land were affected. The area was subjected to a large-scale phyto-management project, and consequently protected as "Green Corridor".
In this study, twenty environmental variables were taken into account and several base models for supervised classification problems were selected, including linear and quadratic discriminant analysis, logistic regression, neural networks and support vector machines. A database with a size of about 30 Mb of alfa-numeric environmental data from the Guadiamar basin was randomly split into three parts, namely training set (50%), validation set (25%), and test set (25%). The techniques were compared from the viewpoint of their accuracy, robustness of results and applicability, and the best models in terms of overall performance were identified. Finally, results were compared with priorities defined in the current regional and national regulations and policies.