EGU23-16375
EGU General Assembly 2023

Machine Learning Models for Spatial Prediction of Groundwater Potentiality in a Large Semi-Arid Mountainous Region: Application to the Rherhaya Watershed, High Atlas, Morocco

Mourad Jadoud[1], Abderrahim El Achheb[1], Noureddine Laftouhi[2], Mustapha Namous[3], **Abdellah Khouz**[4,5,7], Jorge Trindade[4,5,6], Fatima El Bchari[8], Blaid Bougadir[7], Hasna Eloudi[10], and Said Rachidi[9]
[1]Geosciences and Environmental Techniques Laboratory, Faculty of sciences El Jadida, Chouaïb Doukkali University, El Jadida, Morocco
[2]Geosciences Laboratory, Faculty of Sciences Semlalia,Cadi Ayyad University, Marrakesh, Morocco
[3]Data Science for Sustainable Earth Laboratory (Data4Earth), Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco
[4]Universidade Aberta, Lisboa, Portugal
[5]Centre of Geographical Studies, Institute of Geography and Spatial Planning, University of Lisbon, Lisboa, Portugal
[6]Associated Laboratory TERRA, University of Lisbon, Lisboa, Portugal
[7]Higher School of Technology Essaouira, Laboratory of Applied Sciences for the Environment and Sus-tainable Development (SAEDD), Cadi Ayyad University, Essaouira Aljadida, Morocco
[8]Polydisciplinary Faculty of Safi, Department of Earth Sciences, Cadi Ayyad University, Safi, Morocco
[9]Systems Analysis, Information Processing and Industrial Management Laboratory (LASTIMI), High School of Technology-Salé, Mohammadia School of Engineers (EMI), Mohamed V University, Rabat, Morocco
[10]Applied Geology and Geo-Environment Laboratory, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco

Estimating the potential of underground water sources has become a top priority for authorities in semi-arid mountain regions, particularly in the context of recent climate change that has made surface water less accessible and available. However, exploratory methods are typically quite expensive, such as geophysical and topographic methods, which take into account the vastness of the terrain and the extremely limited accessibility in the mountains such as the High Atlas. To achieve this, it has become necessary to use indirect methods first to define potential groundwater zone boundaries before turning to direct measures. This study proposes an indirect, straightforward, quick, and minimally costly method for identifying potential areas for groundwater in the Rherhaya watershed, High Atlas-Morocco, using machine learning, geographical information systems, and remote sensing.

Machine learning algorithms have been increasingly applied to define potential groundwater areas mapping. It's performance to predict the spatial distribution of potential groundwater zones was tested with an inventory-based spring's geodatabase. 254 spring's obtained inventory was split into two independent datasets, including 70% of the springs for the training set and the remaining 30% of springs for validation purposes in the test set. 19 layers of landslide-conditioning factors were prepared and checked for collinearity issues, to produce the potential groundwater zones map. The conditioning factors were selected, prepared and classified, in order to determine the contribution of each class of factors to potential groundwater areas, including: Elevation, Aspect,

Slope angle, Curvature plan, Curvature profile, Stream Power Index (SPI), Topographic Wetness Index (TWI), Normalized Difference Vegetation Index (NDVI), Distance to rivers, Lithology, Rainfall, Land Use and Land Cover (LULC), Drainage density, Valley Depth, Topographic Position Index (TPI), Terrain Ruggedness Index (TRI), Slope Length (LS), Geomorphons, Distance to faults, Distance to spring and Distance to roads. Both the inventory and the conditioning factors are obtained from the observation and interpretation of different data sources, namely high-resolution satellite images, aerial photographs, topographic maps, and extensive field surveys.

Using RStudio software, three machine-learning algorithms were used in this study: Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM). The model's evaluation and validation was assessed using a hybrid approach based on k-cross validation, ROC Curve - AUC, and confusion matrix for the estimation of the predictive performance. The three models provided very encouraging results in terms of identifying the areas that are very likely to produce subsurface water in the Rherhaya basin. The main contributing factors to the potentiality of underground waters, according to the three methods, are the valley depth, TPI, distance to rivers, and curvature plane. With an AUC of 84.4% in the test data, it was clear that the SVM model outperforms the other models for the 70/30 percent subdivision. These findings may contribute to the development of a significant database that will help in the management of water resources in vulnerable areas by decision-makers.