Using data mining methods in forecasting waterspouts over south Aegean Sea, Greece

Nikolaos C. Kamperakis, Panagiotis T. Nastos, and Ioannis T. Matsangouras

Laboratory of Climatology & Atmospheric Environment, Faculty of Geology & Geoenvironment, University of Athens, University Campus, Greece (n.kamperakis@gmail.com)

In this paper an implementation of data mining methods in a database of hazardous atmospheric events, such as waterspout, is presented with the objective to predict their occurrence. There is a need to forecast waterspout development in order to mitigate the impact to the local society, as they significantly affect coastal areas. Data mining process has been implemented in the hazardous atmospheric events database of the Laboratory of Climatology and Atmospheric Environment (LACAE), University of Athens, with respect to the waterspout distribution over south Aegean Sea, Greece. Concerning data mining method, the WEKA (Waikato Environment for Knowledge Analysis), a suite of machine learning software developed by the University of Waikato, New Zealand, has been used in this study under the GNU General Public License.

Several publications during the last years have revealed that south Aegean Sea is a vulnerable area for waterspout formation, as numerous events have caused significant damage when they came onshore (Crete Island). The LACAE database concerns waterspout events that have occurred within the last 9 years (2005-2013). Atmospheric variables from ERA-Interim datasets as well as upper air soundings from Heraklion (LGIR) upper air weather station, such as wind speed/direction and air temperature at specific isobaric levels of middle/low troposphere have been taken into consideration as inputs in data mining process. In addition, the sea surface temperature and several thermodynamic indices have been also used as variables regarding the air-sea interaction and unstable weather conditions during waterspout days, respectively.

Data mining, an interdisciplinary subfield of computer science, adopts algorithms and data analysis tools, involves methods at the intersection of artificial intelligence, machine learning, statistics, and database systems to discover patterns and relationships in data that may be used to make valid predictions.