Geophysical Research Abstracts
Vol. 18, EGU2016-6137, 2016
EGU General Assembly 2016

Recent advances in environmental data mining

Michael Leuenberger and Mikhail Kanevski

University of Lausanne, Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, Lausanne, Switzerland (michael.leuenberger@unil.ch)

Due to the large amount and complexity of data available nowadays in geo- and environmental sciences, we face the need to develop and incorporate more robust and efficient methods for their analysis, modelling and visualization. An important part of these developments deals with an elaboration and application of a contemporary and coherent methodology following the process from data collection to the justification and communication of the results. Recent fundamental progress in machine learning (ML) can considerably contribute to the development of the emerging field - environmental data science.

The present research highlights and investigates the different issues that can occur when dealing with environmental data mining using cutting-edge machine learning algorithms. In particular, the main attention is paid to the description of the self-consistent methodology and two efficient algorithms - Random Forest (RF, Breiman, 2001) and Extreme Learning Machines (ELM, Huang et al., 2006), which recently gained a great popularity. Despite the fact that they are based on two different concepts, i.e. decision trees vs artificial neural networks, they both propose promising results for complex, high dimensional and non-linear data modelling. In addition, the study discusses several important issues of data driven modelling, including feature selection and uncertainties. The approach considered is accompanied by simulated and real data case studies from renewable resources assessment and natural hazards tasks. In conclusion, the current challenges and future developments in statistical environmental data learning are discussed.

References

- Breiman, L., 2001. Random Forests. Machine Learning 45 (1), 5-32.
- Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1-3), 489-501.
- Kanevski, M., Pozdnoukhov, A., Timonin, V., 2009. Machine Learning for Spatial Environmental Data. EPFL Press; Lausanne, Switzerland, p.392.
- Leuenberger, M., Kanevski, M., 2015. Extreme Learning Machines for spatial environmental data. Computers and Geosciences 85, 64-73.