

EGU2020-19060

<https://doi.org/10.5194/egusphere-egu2020-19060>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Predictor dataset selection method for construction of ML-based Models in flood detection using mutual information

Mohammad Mehdi Bateni, Mario Martina, and Luigi Cesarini

University School for Advanced Studies (IUSS), Pavia, Italy (mehdi.bateni@iusspavia.it)

The field of information theory originally developed within the context of communication engineering, deals with the quantification of the information present in a realization of a stochastic process. Mutual information is a measure of mutual dependence between two variables and can be determined from marginal and joint entropies. It is an efficient tool to investigate linear or non-linear dependencies. In this research, some transformed variables, each based on rainfall data from different datasets in Dominican Republic, are adopted in Neural Network and SVM models to classify flood/no-flood events. A selection procedure is used to select skillful inputs to the flood detection model. The relationship between the flood/no-flood output datasets and each predictor (relevance) and also among predictors (redundancy) were assessed based on the mutual information metric. The minimum redundancy between predictors and maximum relevance to the predictand is targeted in order to choose a set of appropriate predictors.