Geophysical Research Abstracts
Vol. 17, EGU2015-3321-1, 2015
EGU General Assembly 2015

Comparison of machine learning algorithms for their applicability in satellite-based optical rainfall retrievals

Hanna Meyer, Meike Kühnlein, Tim Appelhans, and Thomas Nauss

Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Marburg, Germany
(hanna.meyer@geo.uni-marburg.de)

Machine learning (ML) algorithms have been successfully evaluated as valuable tools in satellite-based rainfall retrievals which shows the high potential of ML algorithms when faced with high dimensional and complex data. Moreover, the recent developments in parallel computing with ML offer new possibilities in terms of training and predicting speed and therefore makes their usage in real time systems feasible.
The present study compares four ML algorithms for rainfall area detection and rainfall rate assignment during daytime, night-time and twilight using MSG SEVIRI data over Germany. Satellite-based proxies for cloud top height, cloud top temperature, cloud phase and cloud water path are applied as predictor variables. As machine learning algorithms, random forests (RF), neural networks (NNET), averaged neural networks (AVNNET) and support vector machines (SVM) are chosen.
The comparison is realised in three steps. First, an extensive tuning study is carried out to customise each of the models. Secondly, the models are trained using the optimum values of model parameters found in the tuning study. Finally, the trained models are used to detect rainfall areas and to assign rainfall rates using an independent validation datasets which is compared against ground-based radar data. To train and validate the models, the radar-based RADOLAN RW product from the German Weather Service (DWD) is used which provides area-wide gauge-adjusted hourly precipitation information.
Though the differences in the performance of the algorithms were rather small, NNET and AVNNET have been identified as the most suitable algorithms. On average, they showed the best performance in rainfall area delineation as well as in rainfall rate assignment. The fast computation time of NNET allows to work with large datasets as it is required in remote sensing based rainfall retrievals.
However, since none of the algorithms performed considerably better that the others we conclude that research effort is needed in providing suitable predictors for rainfall rather than in optimizing by the choice of the ML algorithm.