ID: 20180522100359-240
SESSION: 3
PRESENTATION: ORAL

Medclivar 2018 Conference
University of Belgrade
Belgrade, Serbia, 18th to 21st September 2018

Improving spatio-temporal interpolation of daily precipitation using parallelized machine learning and precipitation derived from MSG

M Per?ec Tadi? (1), T Hengl (2)

1) Meteorological and Hydrological Service, Zagreb, Croatia
2) Envirometrix Ltd, Wegeningen, Netherlands

Contact: melita.percec.tadic@cirus.dhz.hr

Preliminary analysis of 1 008 812 daily precipitation records measured on the 572 ground stations during 2005-2010 period in Croatia revealed that the 73% of data are zeros or the precipitation is less than 1 mm, hence data are zero inflated. Most non-zero data belong to [0,1] mm and [1,10] mm class, and just 5% of precipitation amounts are larger. That strongly affects the predictions that underestimate the observations, especially for large precipitation amounts. Spatio-temporal prediction using machine learning algorithm on the daily precipitation data accounts for 55% of the spatio-temporal variability. The machine learning algorithm of random forest RF is implemented through R ranger package. Even though the RF in general does not account for spatial location and auto-correlation which are important in spatio-temporal analysis of precipitation, the attempt is made to include it through covariates and distance fields. The preliminary set of covariates detected that the most influential predictors are time variables cdate representing cumulative time and doy
representing seasonality effect. Than follow annual precipitation CHELSA_precip, altitude, distance to the sea and different buffer distances. Further attempt in improving the results is by taking into account larger data set for the 1981-2016 period, than monthly precipitation grids for Croatia instead of CHELSA precipitation and satellite precipitation data from MSG missions as additional time dependent predictors.