

EGU2020-9063

<https://doi.org/10.5194/egusphere-egu2020-9063>

EGU General Assembly 2020

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



Spatial assessment of inland excess water hazard using combined machine learning and geostatistical methods

Annamária Laborczi¹, Csaba Bozán², Gábor Szatmári¹, János Körösparti², and László Pásztor¹

¹Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Herman Ottó út 15, 1022 Budapest, Hungary (laborczi@rissac.hu)

²Research Department of Irrigation and Water Management, National Agricultural Research and Innovation Centre; Anna Lilyiget 8, 5540 Szarvas, Hungary

Inland excess water (IEW), considered to be a typical Carpathian Basin land degradation problem, is an interrelated natural and human induced phenomenon, which causes several problems in the flat-land regions of Hungary covering nearly half of the country. The term 'inland excess water' refers to the occurrence of inundations outside the flood levee that originate from sources differing from flood overflow, it is surplus surface water forming due to the lack of runoff, insufficient absorption capability of soil or the upwelling of groundwater. There is a multiplicity of definitions, which indicate the complexity of processes that govern this phenomenon. Most of the definitions have a common part, namely, that inland excess water is temporary water inundation that occurs in flat-lands due to both precipitation and groundwater emerging on the surface as substantial sources.

Identification of areas with high risk requires spatial modelling, that is mapping of the specific natural hazard. Various external environmental factors determine the behaviour of the occurrence, frequency of IEW. Spatial auxiliary information representing IEW forming environmental factors were taken into account to support the spatial inference of the locally experienced IEW frequency values. Two hybrid spatial prediction approaches, which combine machine learning and geostatistics, were tested to construct reliable maps, namely regression kriging (RK) and Random Forest with Ordinary Kriging (RFK) using spatially exhaustive auxiliary data on soil, geology, topography, land use and climate. Both methods divides the spatial inference into two parts.

In Regression Kriging the target variable is modelled at first by multiple linear regression (MLR) of the environmental co-variables. Then ordinary kriging is applied on the difference between the reference and the modelled values (residuals). The prediction result map comes from the sum of the MLR model and the interpolated residuals. Random Forest combined with Kriging is a relatively new method applied in digital environmental mapping. In RFK, the deterministic component of spatial variation is modelled by random forest (RF). RF algorithm builds lots of regression trees and the final model relies on averaging the result of the trees, which are grown independently from each other. In RFK the stochastic part of variation is modelled by kriging using the derived residuals. The final map is the sum of the two component predictions.

Comparing the results of the two approaches, we did not find significant differences in their

accuracy in our pilot. However, both methods are appropriate for predicting inland excess water hazard, RFK is suitable for revealing non-linear and more complex relations than RK. Therefore, we suggest the usage of RFK in further predictions and investigations.

Acknowledgement: Our work was supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167).