

EGU2020-22527

<https://doi.org/10.5194/egusphere-egu2020-22527>

EGU General Assembly 2020

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



Making use of open geo-environmental and agricultural datasets to model NO₃ pollution in groundwater bodies

Christian Schneider

Helmholtz-Zentrum für Umweltforschung (UFZ), Leipzig, Germany

In Germany a vast amount of spatial geo-environmental as well as climatic datasets is available. But anthropic data on land-use and agriculture are still very sparse making it difficult to assess the environmental impacts of different agricultural practices. Recently, some data on spatial pattern of crop production as well as livestock production was made publicly available. It opened up the opportunity to model the impact of agriculture on nitrate leaching into groundwater bodies.

A high share of groundwater bodies in Germany contains nitrate levels above the legal threshold of 50 mg l⁻¹. Our study aims to answer the question: to what extent different types of agriculture are contributing to NO₃ leaching into ground water bodies in relation to environmental factors.

We use the random forest (RF) machine learning algorithm to model and predict nitrate exceedance in ground water bodies. The advantage of the RF algorithm is that it has a high predictive accuracy, it is able to use metric as well as multi-level categorical datasets and it calculates a variable importance measure for each predictor used in a model. It therefore gives a measure to which extent each predictor contributes to the accuracy of the model. For this study we applied the RF classification as well as the RF regression algorithms on different spatial scales.

Out of 56 environmental predictor datasets which are of potential importance for NO₃ transport into groundwater bodies 22 were chosen to model NO₃-exceedance.

A recursive variable elimination scheme was applied to calculate minimum predictor sets based on variable importance. In the end the predictor set which resulted in the most accurate NO₃ prediction was identified and used to model groundwater pollution.

RF-modeling proved to be successful on all three scale levels with OBB accuracy between 0.82 and 0.95. At all scale levels environmental co-variables played a major role in predicting NO₃-exceedance. But the RF variable importance measure could also be used to identify the contribution of agricultural predictors to NO₃ exceedance and to quantitatively proof our hypotheses.

One main challenge was to identify the influence of data quality on the RF variable importance

measure.