

EGU21-9704

<https://doi.org/10.5194/egusphere-egu21-9704>

EGU General Assembly 2021

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



Measurement error-filtered machine learning in digital soil mapping

Stephan van der Westhuizen^{1,2}, Gerard Heuvelink^{2,3}, and David Hofmeyr¹

¹Stellenbosch University, Stellenbosch, South Africa

²Wageningen University and Research, Wageningen, The Netherlands

³ISRIC - World Soil Information, Wageningen, The Netherlands

Digital soil mapping (DSM) may be defined as the use of a statistical model to quantify the relationship between a certain observed soil property at various geographic locations, and a collection of environmental covariates, and then using this relationship to predict the soil property at locations where the property was not measured. It is also important to quantify the uncertainty with regards to prediction of these soil maps. An important source of uncertainty in DSM is measurement error which is considered as the difference between a measured and true value of a soil property.

The use of machine learning (ML) models such as random forests (RF) has become a popular trend in DSM. This is because ML models tend to be capable of accommodating highly non-linear relationships between the soil property and covariates. However, it is not clear how to incorporate measurement error into ML models. In this presentation we will discuss how to incorporate measurement error into some popular ML models, starting with incorporating weights into the objective function of ML models that implicitly assume a Gaussian error. We will discuss the effect that these modifications have on prediction accuracy, with reference to simulation studies.