Comparison of uncertainty quantification methods on the example of soil organic carbon (SOC) stock mapping in Hungary

Gábor Szatmári

Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Budapest, Hungary

László Pásztor

Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Budapest, Hungary

Research Highlights

- The applied digital soil mapping (DSM) techniques provided different uncertainty models with different performances.
- In point of uncertainty quantification, sequential Gaussian simulation and quantile regression forest outperformed the others.
- We have demonstrated that uncertainty models must be validated.
- Special attention should be paid to the assumptions made in uncertainty modelling.

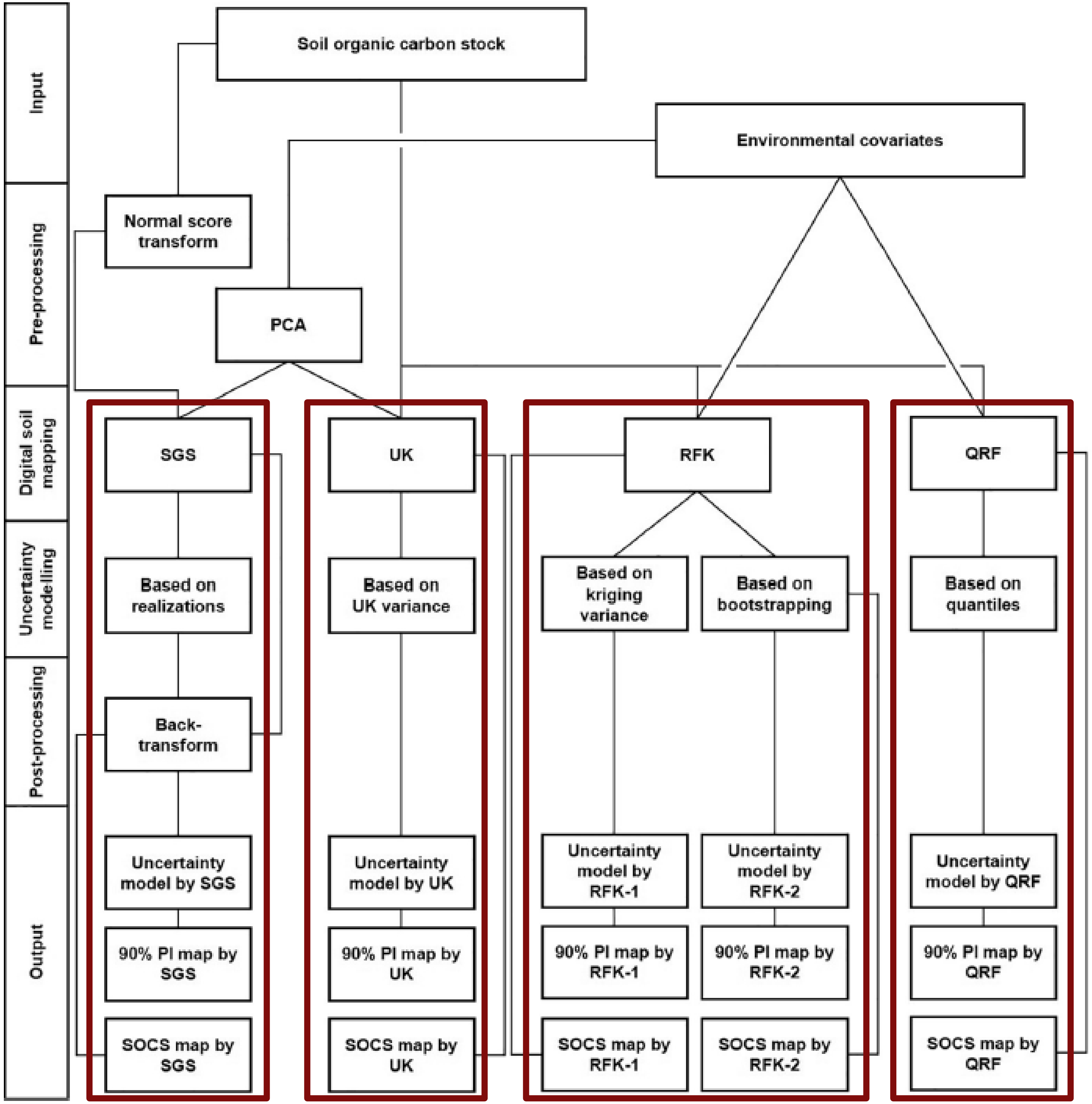
Flowchart

Applied DSM techniques:

1. Sequential Gaussian simulation (SGS)
2. Universal kriging (UK)
3. Random forest + kriging (RFK)
   3.1. Based on kriging variance (RFK-1)
   3.2. Based on bootstrapping (RFK-2)
4. Quantile regression forest (QRF)

Abbreviations:
SOCS: soil organic carbon stock,
PCA: principal component analysis,
PI: prediction interval

Source: Szatmári & Pásztor (2019)
https://doi.org/10.1016/j.geoderma.2018.09.008



SOC stock maps & their uncertainties

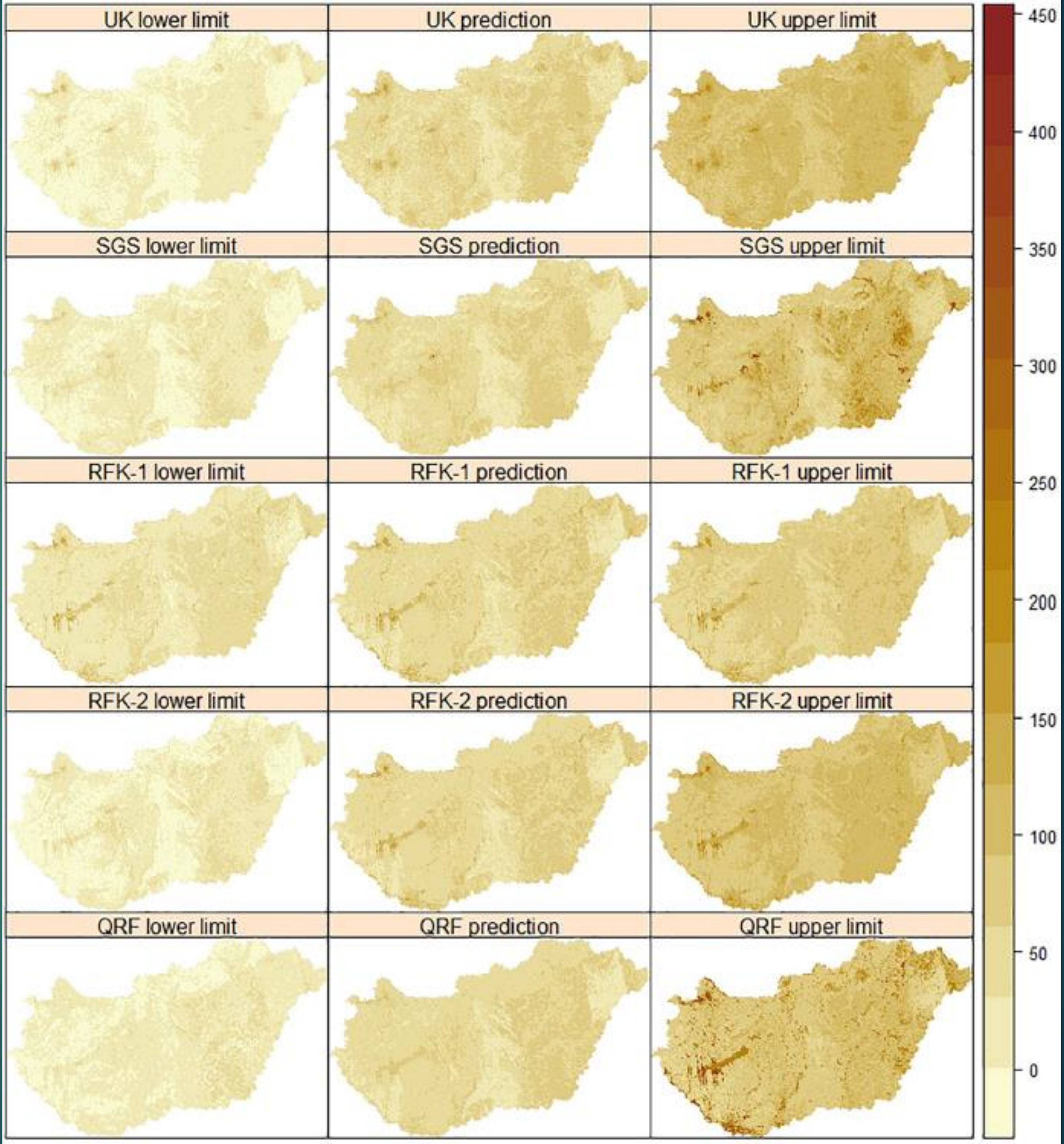
Visualization of uncertainty:

The upper and lower limit of the 90% prediction interval are presented. This prediction interval reports the range of values within which the true value is expected to occur 9 times out of 10.

The unit of the maps is **tons · ha^{-1}**

The geometric resolution of the maps is **500 m**

Source: Szatmári & Pásztor (2019)
https://doi.org/10.1016/j.geoderma.2018.09.008



Validation of uncertainty quantifications

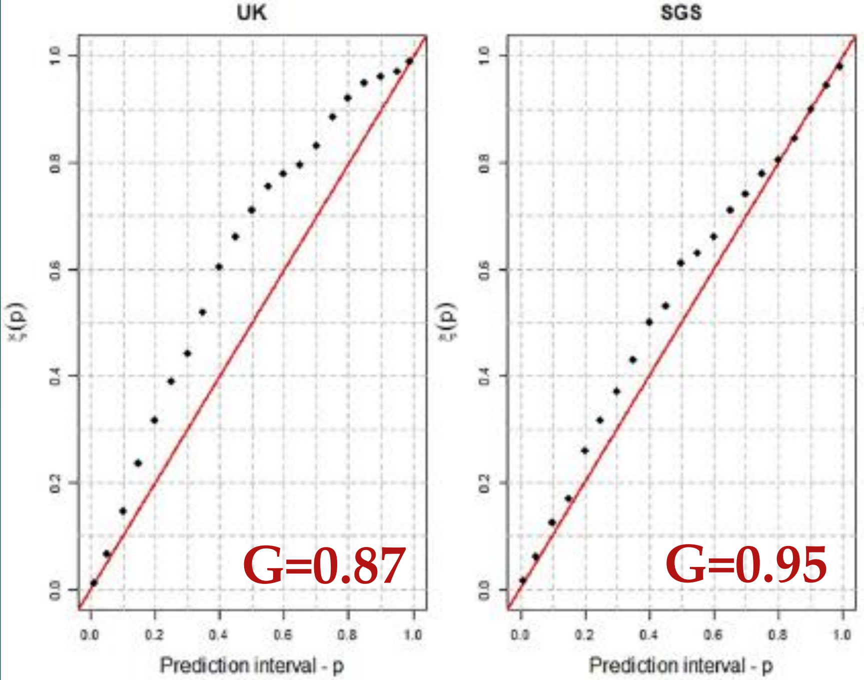
Validation was carried out by using 200 independent SOC stock observations

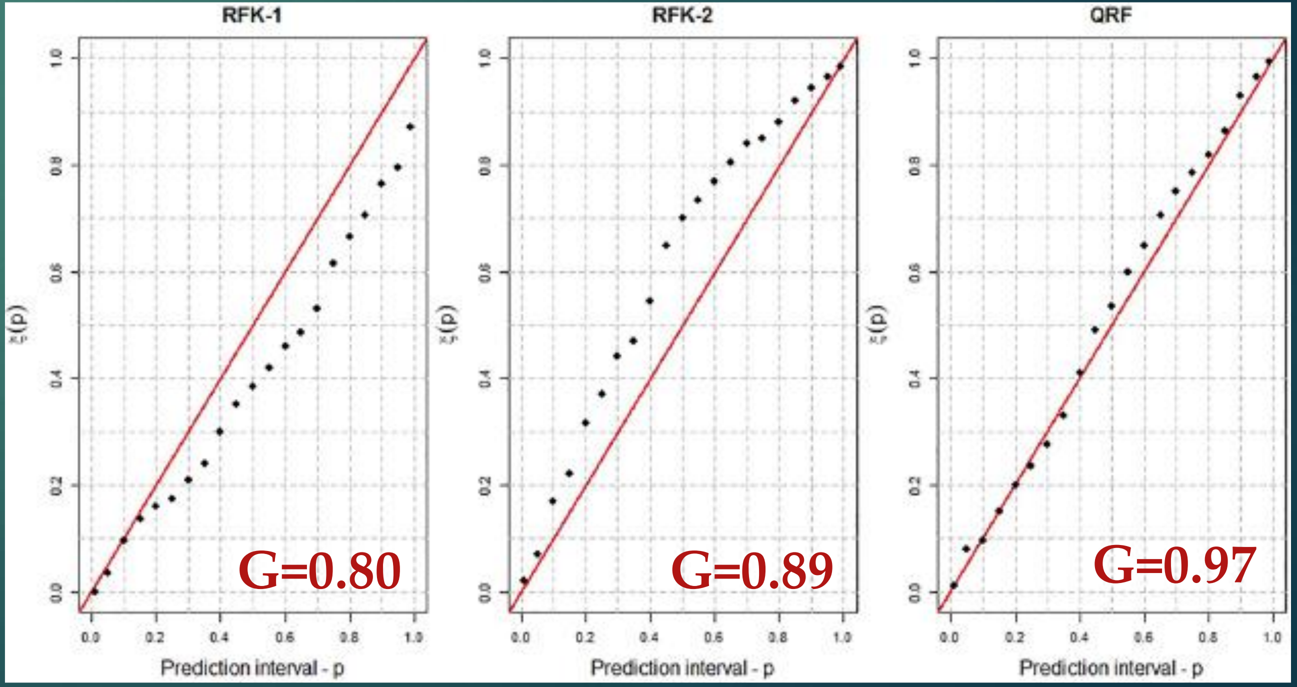
<u>Rank of the applied DSM techniques (from best to worst):</u>

QRF, SGS, RFK-2, UK, RFK-1

Accuracy plot graphically shows the actual fraction of true values falling within symmetric prediction intervals of varying width.

The value of G shows the closeness of the actual and expected fractions. Ideally, G is equal to 1.

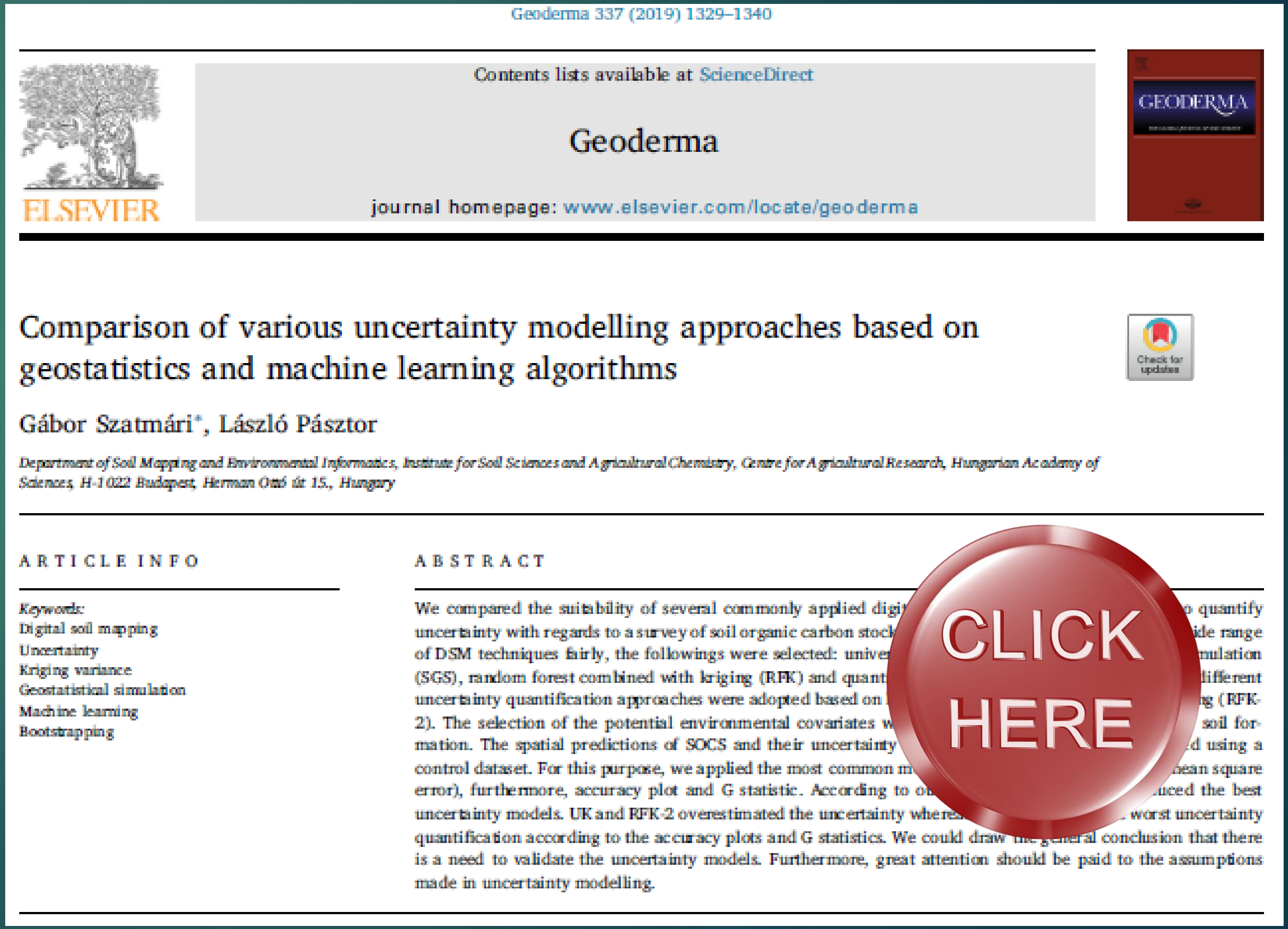




Source: Szatmári & Pásztor (2019)
https://doi.org/10.1016/j.geoderma.2018.09.008

If you want to know more…

…You can take a look at the paper below



Geoderma 337 (2019) 1329–1340

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms

Gábor Szatmári*, László Pásztor

Department of Soil Mapping and Environmental Informatics, Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, H-1022 Budapest, Herman Ottó út 15., Hungary

ARTICLE INFO

Keywords:
Digital soil mapping
Uncertainty
Kriging variance
Geostatistical simulation
Machine learning
Bootstrapping

ABSTRACT

We compared the suitability of several commonly applied digi … o quantify uncertainty with regards to a survey of soil organic carbon stock … ide range of DSM techniques fairly, the followings were selected: unive … mulation (SGS), random forest combined with kriging (RFK) and quant … different uncertainty quantification approaches were adopted based on … g (RFK-2). The selection of the potential environmental covariates w … soil formation. The spatial predictions of SOCS and their uncertainty … d using a control dataset. For this purpose, we applied the most common m … ean square error), furthermore, accuracy plot and G statistic. According to o … uced the best uncertainty models. UK and RFK-2 overestimated the uncertainty where … worst uncertainty quantification according to the accuracy plots and G statistics. We could draw the general conclusion that there is a need to validate the uncertainty models. Furthermore, great attention should be paid to the assumptions made in uncertainty modelling.

OPEN ACCESS

https://doi.org/10.1016/j.geoderma.2018.09.008