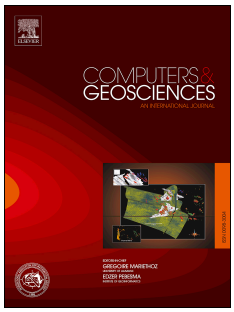
Accepted Manuscript



Sparse regression interaction models for spatial prediction of soil properties in 3D

Milutin Pejović, Mladen Nikolić, Gerard B.M. Heuvelink, Tomislav Hengl, Milan Kilibarda, Branislav Bajat

PII: S0098-3004(17)30852-X

DOI: 10.1016/j.cageo.2018.05.008

Reference: CAGEO 4133

To appear in: *Computers and Geosciences*

Received Date: 8 August 2017

Revised Date: 29 March 2018

Accepted Date: 8 May 2018

Please cite this article as: Pejović, M., Nikolić, M., Heuvelink, G.B.M., Hengl, T., Kilibarda, M., Bajat, B., Sparse regression interaction models for spatial prediction of soil properties in 3D, *Computers and Geosciences* (2018), doi: 10.1016/j.cageo.2018.05.008.

Sparse Regression Interaction Models for Spatial Prediction of Soil Properties in 3D

Milutin Pejović[a], Mladen Nikolić[b], Gerard B.M. Heuvelink[c], Tomislav Hengl[d], Milan Kilibarda[a], Branislav Bajat[a,*]

[a]*Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Belgrade, Serbia*
[b]*Faculty of Mathematics, University of Belgrade, Studentski Trg 16, 11000 Belgrade, Serbia*
[c]*ISRIC — World Soil Information, Wageningen, The Netherlands*
[d]*Envirometrix Ltd., Wageningen, the Netherlands*

Abstract

An approach for using lasso (Least Absolute Shrinkage and Selection Operator) regression in creating sparse 3D models of soil properties for spatial prediction at multiple depths is presented. Modeling soil properties in 3D benefits from interactions of spatial predictors with soil depth and its polynomial expansion, which yields a large number of model variables (and corresponding model parameters). Lasso is able to perform variable selection, hence reducing the number of model parameters and making the model more easily interpretable. This also prevents overfitting, which makes the model more accurate. The presented approach was tested using four variable selection approaches – none, stepwise, lasso and hierarchical lasso, on four kinds of models – standard linear model, linear model with polynomial expansion of depth, linear model with interactions of covariates with depth and linear model with interactions of covariates with depth and its polynomial expansion. This framework was used to predict Soil Organic Carbon (SOC) in three contrasting study areas: Bor (Serbia), Edgeroi (Australia) and the Netherlands. Results show that lasso yields substantial improvements in accuracy over standard and stepwise regression — up to 50 % of total variance. It yields models which contain up to five times less nonzero parameters than the full models and that are usually more sparse than models obtained by stepwise regres-

*Corresponding author: Branislav Bajat, University of Belgrade, Faculty of Civil Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia Tel.: +381-(0)11-3218579

Email addresses: mpejovic@grf.bg.ac.rs (Milutin Pejović), nikolic@matf.bg.ac.rs (Mladen Nikolić), gerard.heuvelink@wur.nl (Gerard B.M. Heuvelink), tom.hengl@envirometrix.net (Tomislav Hengl), kili@grf.bg.ac.rs (Milan Kilibarda), bajat@grf.bg.ac.rs (Branislav Bajat)