Title: An Insight into Machine Learning AlgorithmstoMapthe Occurrence of Soil MatticHorizon in the Northeastern Qinghai-Tibetan Plateau

Author: ZHI Junjun, ZHANG Ganlin, YANGRenmin, YANGFei, JINChengwei, LIUFeng, SONGXiaodong, ZHAOYuguo, LI Decheng

PII: S1002-0160(17)60481-8
DOI: 10.1016/S1002-0160(17)60481-8
Reference: NA

To appear in:

Received date: NA
Revised date: NA
Accepted date: NA

Please cite this article as: ZHI Junjun, ZHANG Ganlin, YANGRenmin, YANGFei, JINChengwei, LIUFeng, SONGXiaodong, ZHAOYuguo, LI Decheng, An Insight into Machine Learning AlgorithmstoMapthe Occurrence of Soil MatticHorizon in the Northeastern Qinghai-Tibetan Plateau, *Pedosphere* (2017), 10.1016/S1002-0160(17)60481-8.

PEDOSPHERE
Pedosphere ISSN 1002-0160/CN 32-1315/P doi:10.1016/S1002-0160(17)60481-8

An Insight into Machine Learning AlgorithmstoMapthe Occurrence of Soil MatticHorizon in the Northeastern Qinghai-Tibetan Plateau

ZHI Junjun[1], ZHANG Ganlin[1,2,*], YANGRenmin [1,2], YANGFei[1,2], JINChengwei [1], LIUFeng[1], SONGXiaodong [1], ZHAOYuguo[1], LI Decheng [1]

[1]*State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008(China)*

[2]*University of the Chinese Academy of Sciences, Beijing 100049(China)*

ABSTRACT

Soil diagnostic horizons, which havea set of quantified properties,play a key role in soil classification. However, they are difficult to predict, and little attempt has been made to maptheir spatial occurrence.We evaluated and compared four machine learning algorithms, namely classification and regression tree (CART), random forest (RF), boosted regression trees (BRT), and support vector machine (SVM),to map the occurrence of soil mattichorizon in the northeastern Qinghai-Tibetan Plateau using easily accessible ancillary data.The mechanism of resampling and ensemble techniques significantly improved prediction accuracies (measured by area under the Receiver Operator Characteristic curve score, AUC),and produced more stable results of the BRT (AUC of 0.921 ± 0.012, mean ± standard deviation) and RF (0.908 ± 0.013) models compared with CART (0.784 ± 0.012) which isthe most commonly used machine learning method.Although the SVM model yielded a comparable AUC value (0.906 ± 0.006) to RF and BRT, it is sensitive to parameter settings which are extremely time-consuming, thus we consider it inadequate for occurrence-distribution modeling.Consideringthe obvious advantages in high prediction accuracy, robust to parameter settings, the ability to estimate uncertainty in prediction, and easy interpretation of predictor variables, BRTseems to be the most adequate method. The outcomes provide an insight into the use of machine learning algorithms to map the mattic horizon, and perhaps others.

KeyWords: digital soil mapping; machine learning algorithm; Qinghai-Tibetan Plateau; soil diagnostic horizons

INTRODUCTION

Soil diagnostic horizons provide baseline information for soil classification, agricultural activities, and ecological management.On the Qinghai-Tibetan Plateau, lowtemperaturesand water stagnation in the

*Corresponding author. E-mail:glzhang@issas.ac.cn.