A two point machine learning method for spatial prediction for soil: overcoming the spatially heterogeneous distribution and relationship of soil heavy metal concentration

Gao Bingbo¹, Alfred Stein², and Wang Jinfeng³

¹College of land science and technology, China agricultural university, Beijing, China
²Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
³Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing, China

The soil heavy metal contamination has becoming a serious problem worldwide. An accurate prediction of soil heavy metal concentration at un-sampled locations using a small sample remains a challenge, because of many natural and human factors and resulted complex heterogeneous pattern, and the relationship between influencing factors are also not homogeneous. To overcome those heterogeneities and improve the prediction accuracy, a two point machine learning method is proposed in this paper by fully leveraging the spatial relationship and similarity relationship of high dimensional ancillary variables. It firstly models the difference between paired points using machine learning model, then predict the concentration differences between sampling points and the un-sampled points, and finally utilize the predicted differences to choose near neighbors to obtain the final concentration prediction. In this method, an innovative way to search near neighbors for local model from the difference of response variable was put forward to overcome the Curse of Dimensionality. Its performance was illustrated in two diverse case studies and it is demonstrated that proposed method can dramatically improve the prediction accuracy for soil heavy metal. Besides spatial prediction of soil pollution, it can also be applied to spatial prediction of other other elements of the earth system. And in further the machine learning method in this paper can be replaced to any other supervised learning model according to specific situations.