Hyperspectral Detection of Soil Organic Matter Content Based on Random Forest Algorithm: Postprint

Authors: Bao Qingling, Ding Jianli, Wang Jingzhe, Cai Lianghong, Ding Jianli

Date: 2019-11-15T00:00:00+00:00

Abstract

To explore rapid detection of soil organic matter content that both preserves spectral information and ensures accuracy, 73 soil sampling points and their corresponding hyperspectral data from within the Weigan River-Kuqa Oasis in southern Xinjiang were used as the research subject. Wavelet transform and mathematical transformation methods were applied for spectral data preprocessing to analyze differences in spectral curves of various wavelet decomposition and reconstruction spectra under different organic matter contents and soil types, to determine the maximum wavelet decomposition level and screen sensitive bands through correlation analysis, to conduct importance analysis on feature spectra from each wavelet decomposition level by combining grey relational analysis with random forest prediction classification models, and finally to establish and analyze a multiple linear prediction model based on optimal feature spectra. The results show that: (1) The spectral curves of cultivated soil and forest soil exhibit relatively gentle variations compared to those of saline soil and desert soil, while at water absorption bands, the spectral curves of saline soil show the deepest absorption valleys. (2) The correlation between wavelet transform decomposition spectra and soil organic matter content displays a trend of first decreasing then increasing with the increase of decomposition levels; at the 6th level, the variations in characteristic spectral curves and the number of sensitive bands tend to stabilize, which is determined as the maximum decomposition level for wavelet transform. (3) Compared with grey relational analysis, the random forest model meets expectations for screening factors at each wavelet decomposition level, with the ranking of influence on soil organic matter content from high to low being L3-(1/LgR) , L4-(1/LgR) , L6-(1/LgR) , L5-(1/LgR) , L2-(1/LgR) , L0-1/LgR, L1-1/LgR. (4) In wavelet decomposition spectra, the mid-frequency range feature spectra demonstrate superior estimation capability for soil organic matter content in arid regions compared to high-frequency and low-frequency range feature spectra, while the model established based on

L-MC achieves the highest accuracy. The study demonstrates that soil spectral organic matter content monitoring based on machine learning classification methods combined with wavelet decomposition can effectively reduce noise band interference and improve the classification and prediction accuracy of characteristic bands.

Full Text

Hyperspectral Detection of Soil Organic Matter Content Based on Random Forest Algorithm

BAO Qing-ling[1,2], DING Jian-li[1,2], WANG Jing-zhe[1,2], CAI Liang-hong[1,2]

[1]Key Laboratory of Wisdom City and Environmental Modeling, Department of Education, Xinjiang University, Urumqi 830046, Xinjiang, China
[2]Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, Xinjiang, China

Abstract

To explore methods for retaining spectral information while accurately detecting soil organic matter content, this study investigated the potential of spectral processing techniques—specifically wavelet decomposition and random forest methods—for estimating soil organic matter content and analyzing the spectral curves of various wavelet decomposition reconstruction spectra across different soil types. The research focused on soil samples collected from the Weigan River Oasis in Kuqa County, a typical arid-region oasis in the northern Tarim Basin of Xinjiang, China. Soil organic matter content was determined for these samples, and their spectra were measured using an ASD FieldSpecFR instrument. The spectral data were then preprocessed through wavelet decomposition and mathematical transformation. Discrete Wavelet Transform (DWT), with its multi-scale analysis capability, was applied to decompose soil near-infrared spectroscopy data at multiple scales to analyze wavelet decomposition reconstruction spectral curves under varying organic matter contents and soil types. Correlation analysis was employed to determine the optimal wavelet decomposition layer and sensitive bands. Finally, a multivariate linear prediction model for soil organic matter content was established based on the optimal characteristic spectrum, which was identified by combining grey correlation analysis and random forest methods to evaluate the significance of different wavelet decomposition characteristic spectra. The results demonstrated: (1) The spectral reflectance of each wavelet decomposition decreased with increasing organic matter content, while the spectral curves of cultivated and forest soils exhibited more gradual changes compared to saline-alkali and desert soils. (2) The correlation between the wavelet transform decomposition spectrum and soil organic matter content initially decreased then increased with additional decomposition layers; in the sixth layer, the characteristic spectral curve and number of sensitive bands stabilized, suggesting this layer as the optimal maximum decomposition level. (3) Compared with grey correlation analysis, the random forest model more effectively screened factors at each wavelet decomposition layer, ranking them by impact on soil organic matter content in descending order as follows: L3-(1/LgR) ', L4-(1/LgR)' , L6-(1/LgR) ', L5-(1/LgR)' , L2-(1/LgR)' , L0-1/LgR, L1-1/LgR. (4) Statistical analysis of all SOM estimation models indicated that the L-MC-based model achieved the highest accuracy. This research demonstrates that monitoring soil spectral organic matter content using machine learning classification combined with wavelet decomposition can effectively reduce noise band interference and improve classification prediction accuracy for characteristic bands. The random forest prediction model offers significant advantages over traditional linear prediction models such as grey correlation analysis, showing superior statistical performance, reliability, and predictive stability. These results provide valuable scientific reference and support for soil nutrient research in arid regions and local precision agriculture.

Keywords: hyperspectral; soil organic matter content; wavelet transform; random forest

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.