IECAG 2023

Temperature-dependent IR Spectroscopy of Silicate Soils with Cluster Analysis for Soil Differentiation

RSF Russian Science Foundation 19-13-00117

Mikhail A. Proskurnin[1,*], Dmitry S. Volkov[1,2], Dmitry M. Filatov[1], Ivan V. Mikheev[1], Olga B. Rogova[1,2]

[1] Chemistry Department, M.V. Lomonosov Moscow State University, Leninskie Gory, 1-3, 119991, Moscow, Russia, proskurnin@gmail.com (M.A.P.); dim020202@mail.ru (D.M.F.)

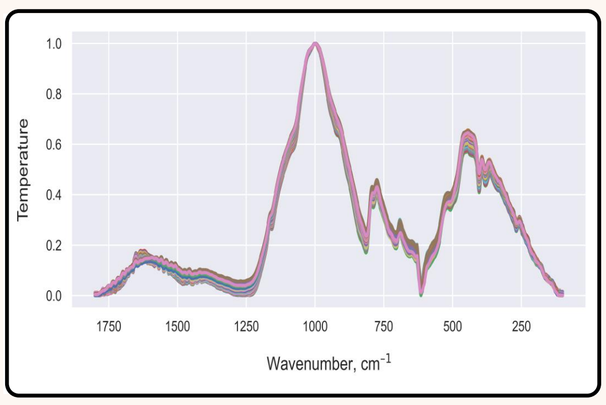
[2] Department of Chemistry and Physical Chemistry of Soils, V.V. Dokuchaev Soil Science Institute, Pyzhevsky per., 7/2, 119017, Moscow, Russia, obrogova@gmail.com (O.B.R.)

Introduction

Temperature-dependent infrared spectroscopy (TIR) is a highly informative technique for complex samples that can shed light on the problem of detecting differences in soil properties both at the level of soil organic matter composition and changes in the silicate matrix part. In combination with machine-learning (ML)-based dimensionality reduction methods, it was used to develop a model for cluster analysis of silicate matrix soils in terms of soil organic matter (SOM), SOM-matrix interactions and biogenic silica, and this approach was compared with classical analysis approaches.

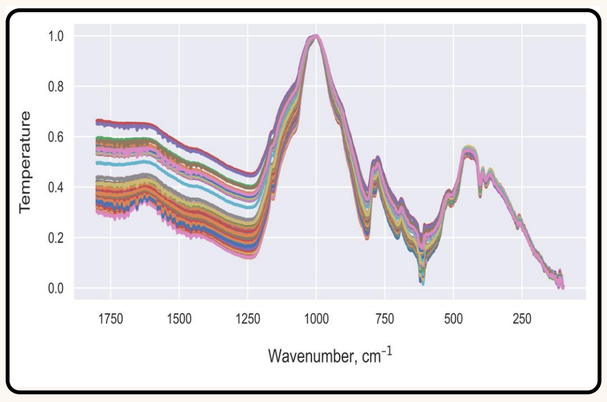
Agricultural land-use chernozems (native, fallow, cropland, and shelterbelt) were subjected to granulometric fractionation to obtain a broad range of soil particles and microaggregates. Silt (size, <2 μm), dust (2–5 μm), cutoff (<20 and <50 μm), and narrow (20–30, 40–50, 50–100, and 100–200 μm) fractions were selected as characteristic.

Results: ML-based Approach for Cluster Analysis of Soils



Temperature-dependent spectrum Steppe (50-100 μm), 1800-100 cm^{-1}

Temperature-dependent spectrum Fallow (50-100 μm), 1800-100 cm^{-1}

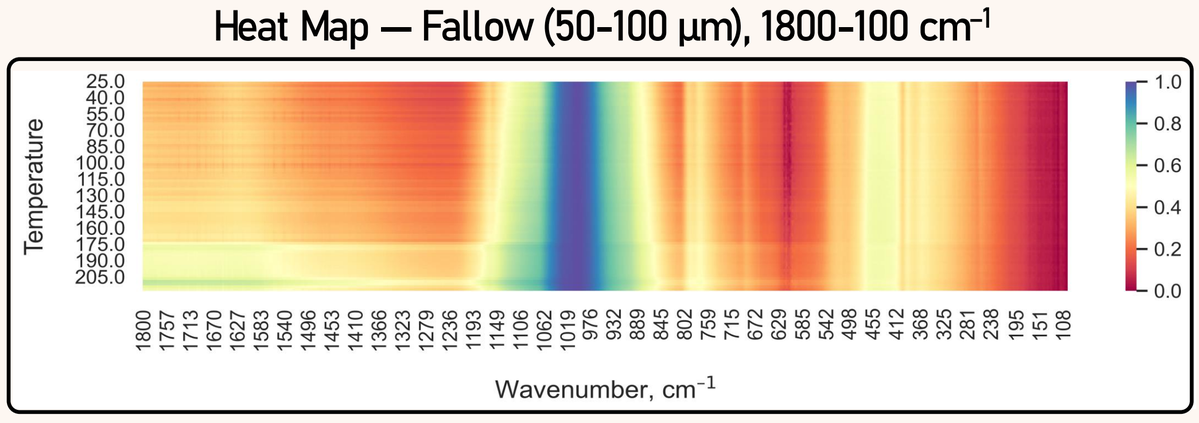


(1) Different thermal behavior of bands assigned to SOM and inorganic matrix was found. Among them are hydrogen-bond region (4000–3200 cm^{-1}), aromatic/aliphatic fragments (3100–2800 cm^{-1}), carboxylic acids, carboxylates, and other SOM functional groups (1800–1200 cm^{-1}), and bands associated with phytoliths and quartz (850–150 cm^{-1}). Analysis of the individual temperature-dependent spectra is the simplest approach, but it cannot provide accurate cluster structure estimation.

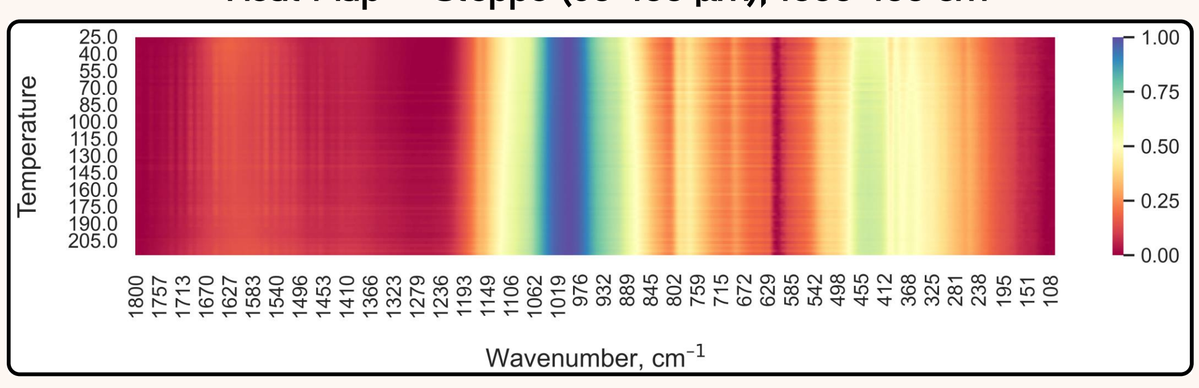
(2) Heat map TIR analysis provides a more detailed assessment of spectral similarities than their visual assessment. Changes in band frequencies and integral intensities in TIR separates land-use samples based on the vegetation and agrogenic loads. Main contributions are changes in biogenic quartz and phytoliths. However, TIR does not provide a clear explanation of the cluster structure. It also cannot distinguish fine fractions (dust, silt, and <20 μm).

Approach to Cluster Analysis of Temperature-Dependent IR spectra = Scaling to range [0...1] + Heat map plot + PCA & t-SNE cluster maps + Total Analysis

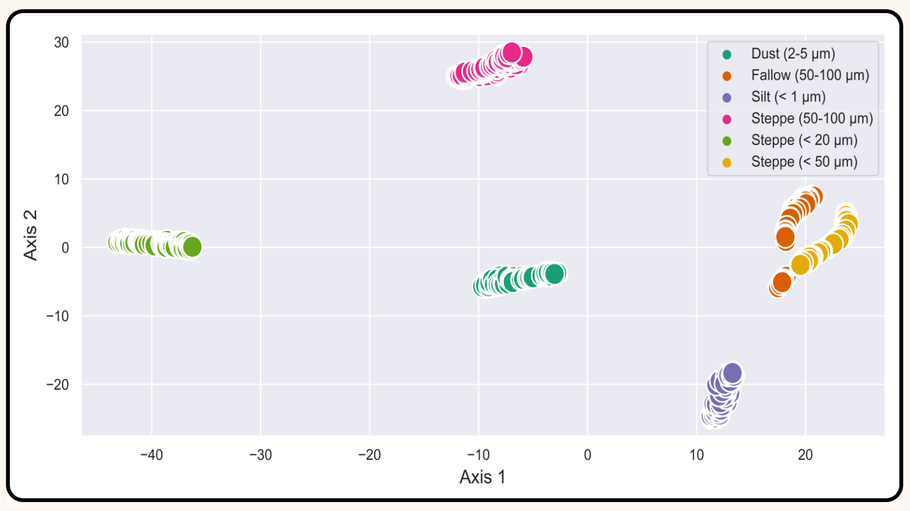
Heat Map — Fallow (50-100 μm), 1800-100 cm^{-1}



Heat Map — Steppe (50-100 μm), 1800-100 cm^{-1}



t-SNE Cluster Map, 1800-100 cm^{-1}



(1) The use of dimensionality reduction techniques allows for the best estimation of the cluster structure. By mapping the original spectrum into a point on a two-dimensional plane, we get rid of insignificant signals and noise and obtain a clear visualization, based on which we can easily judge whether a given soil type is separated from others or not.

(2) The best result was achieved at ranges of 1800–100 cm^{-1} (SOM and matrix together) and 850–150 cm^{-1} (SiO_2 range) by the t-SNE method, which uses a nonlinear transformation to reduce the dimensionality of the data.

(3) Determination of the most informative ranges of the spectra was carried out by means of a joint analysis of the scatter plots obtained by t-SNE method and heat maps, analysis of heat maps only does not provide a reliable estimation.

(4) Thus, ML-TIR-based t-SNE approach provided the separation of fine-size samples with increased contributions from SOM due to dimensionality reduction.

Example: Soil Samples Differentiation Results & Key Spectral Ranges for the Differentiation Basis

Basis of Approach to Cluster Analysis	Silt (<1 μm)	Dust (2-5 μm)	Steppe (<20 μm)	Steppe (<50 μm)	Steppe (50-100 μm)	Fallow (50-100 μm)
Temperature Spectra	✓/X	✓/X	✓	✓/X	✓/X	✓/X
Heat map of Spectra	✓	✓/X	✓/X	✓/X	✓/X	✓/X
Cluster map (t-SNE) 1800-100 cm^{-1}	✓	✓	✓	✓/X	✓	✓/X
Cluster map (t-SNE) 850-150 cm^{-1}	✓	✓	✓	✓/X	✓	✓/X

1800–100 cm^{-1} ↔ 1800-1250 cm^{-1} | 450 cm^{-1} | 350 cm^{-1}

850–150 cm^{-1} ↔ 850-550 cm^{-1} | 780-750 cm^{-1} | 390-350 cm^{-1}

3. Conclusions

Comparison of approaches to cluster analysis of temperature-dependent IR spectral data of soils shown that the analysis of data of reduced dimensionality is the best from the viewpoint of information conext; it allows more clearly determining the differences in samples associated with differences in spectra due to both the size and chemical composition.