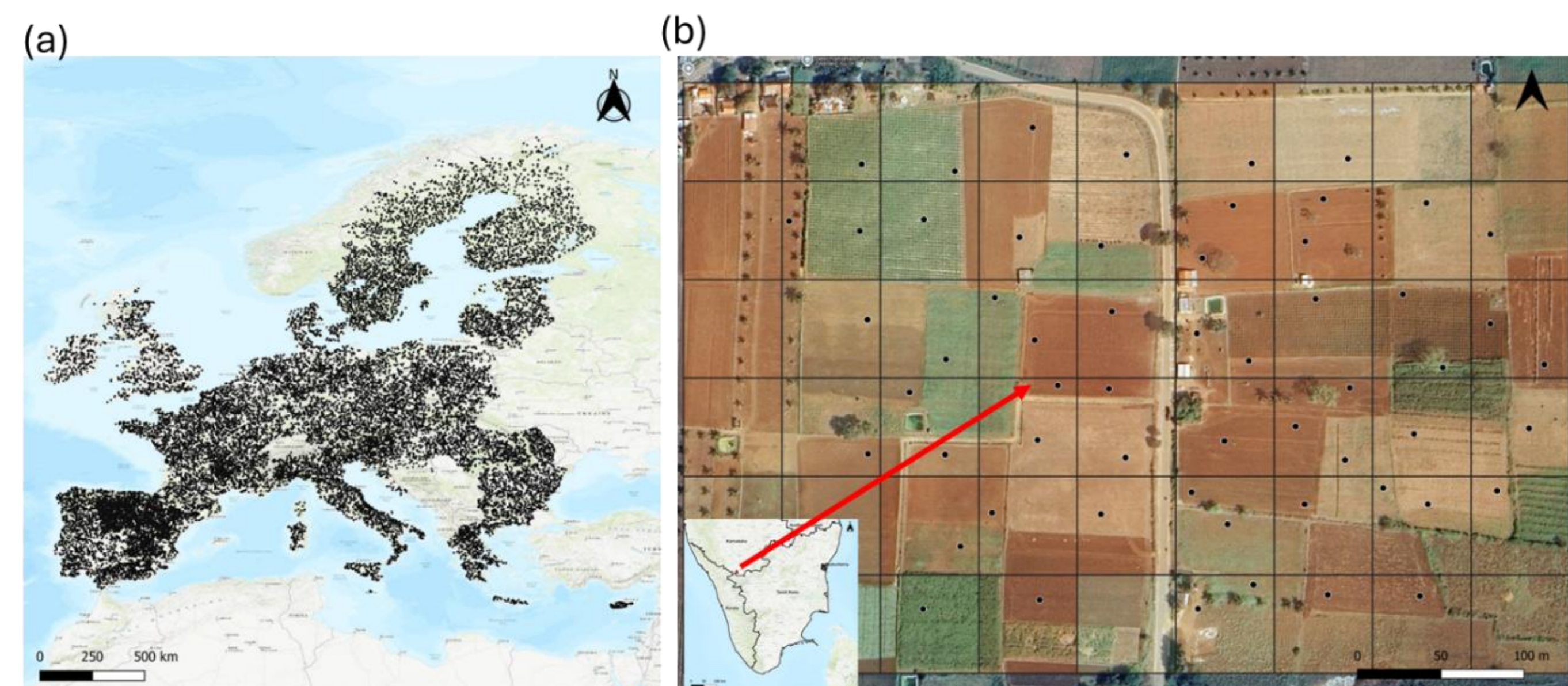


Motivation

- Soil property estimation is essential for environmental monitoring, sustainable land management, and climate modeling.
- Hyperspectral data offers new opportunities, but differences in sensors and acquisition conditions, limit model transferability across regions and platforms.
- Spectral preprocessing improves cross-sensor consistency, enabling more robust and scalable soil property models from multi-source spectral data.

Data

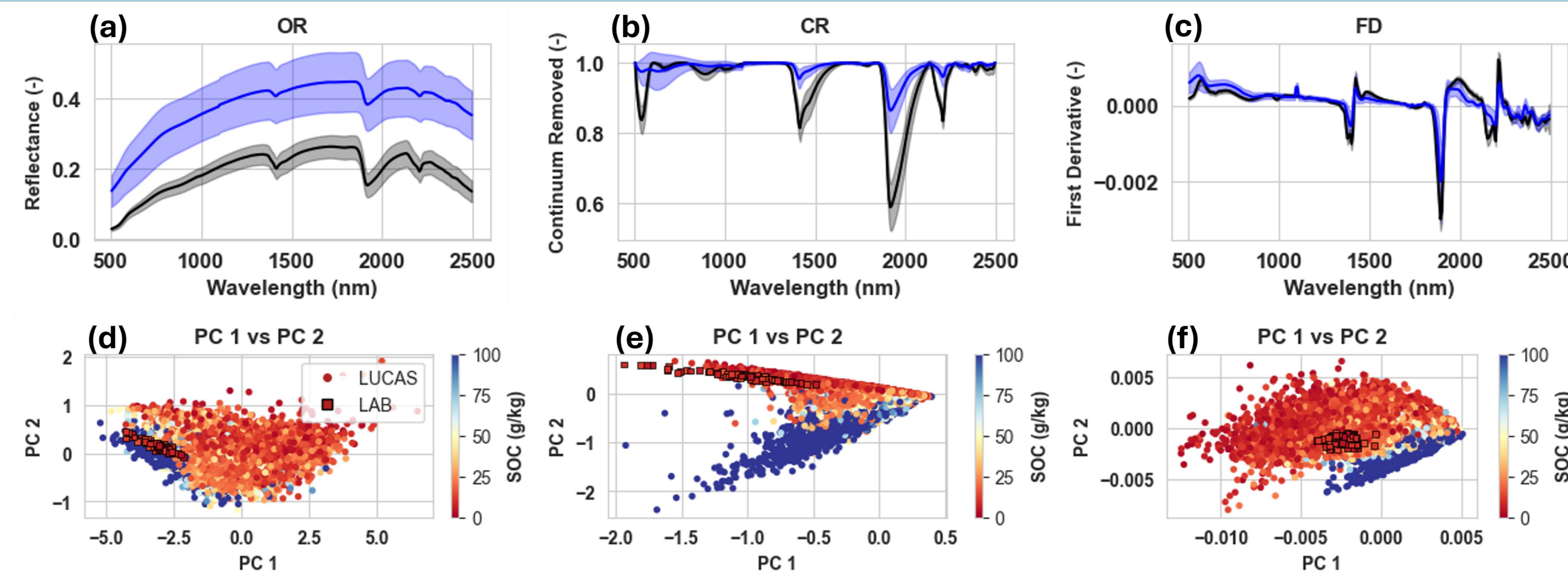


(a) Soil sampling locations across Europe from the LUCAS dataset. (b) Detailed view of sampling locations within agricultural fields in Berambadi, Karnataka, India.

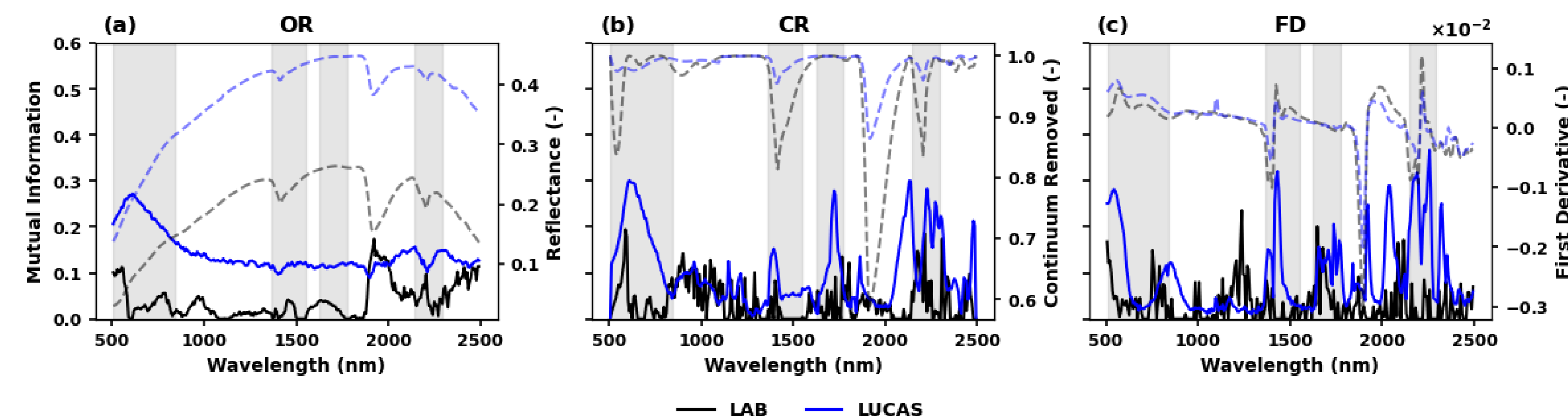
Methods

- Preprocessed LAB and LUCAS soil spectra
 - Original Reflectance (OR)
 - Continuum Removal (CR)
 - First Derivative (FD)
- Estimated Soil organic carbon (SOC) using LASSO, PLSR, and RF models, with stratified analysis using mutual information.
- Continuum Removal Feature Analysis

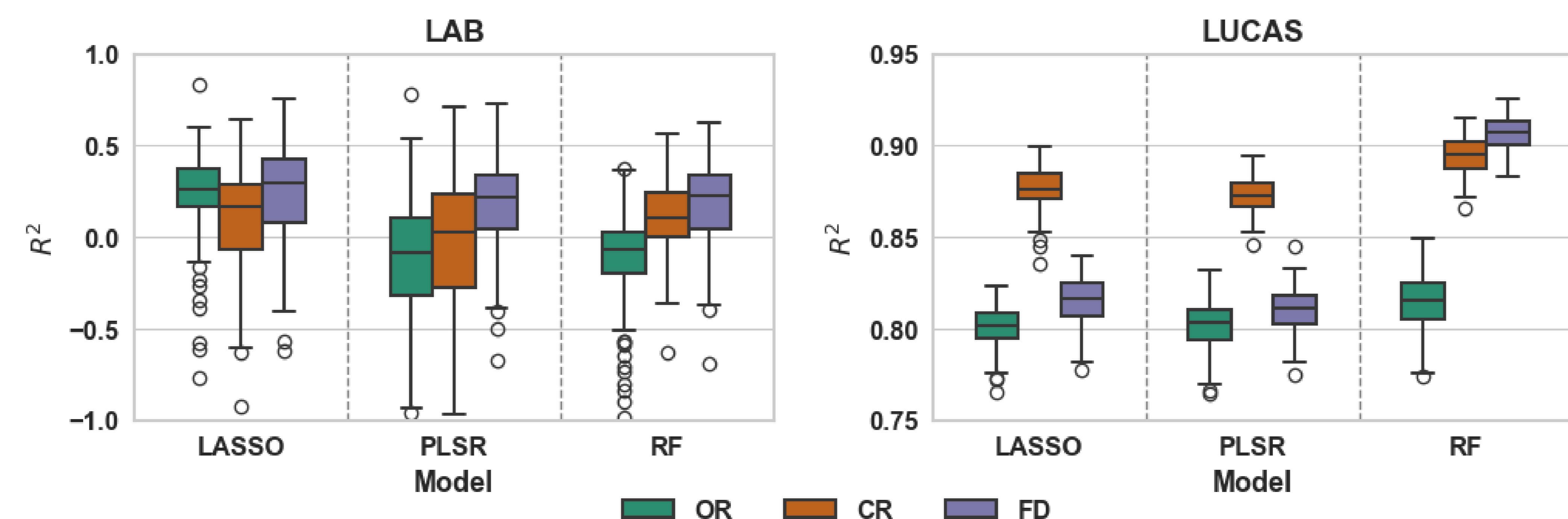
Results



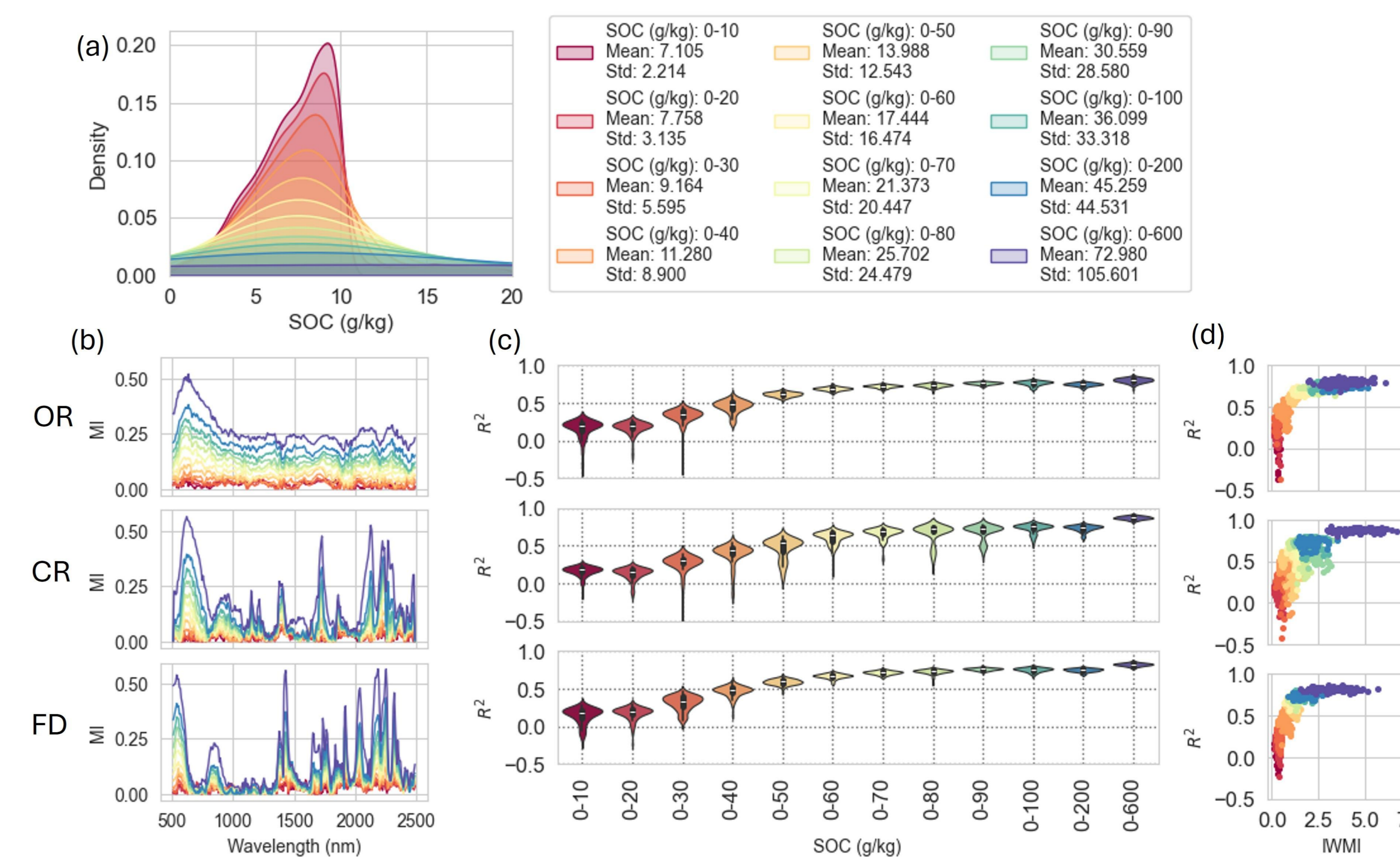
Spectral characteristics and SOC distribution. (a–c) Mean and standard deviation of OR, CR, and FD soil spectra. (d–f) Principal component biplots (PC1 vs. PC2) of OR, CR, and FD spectra, colored by SOC values.



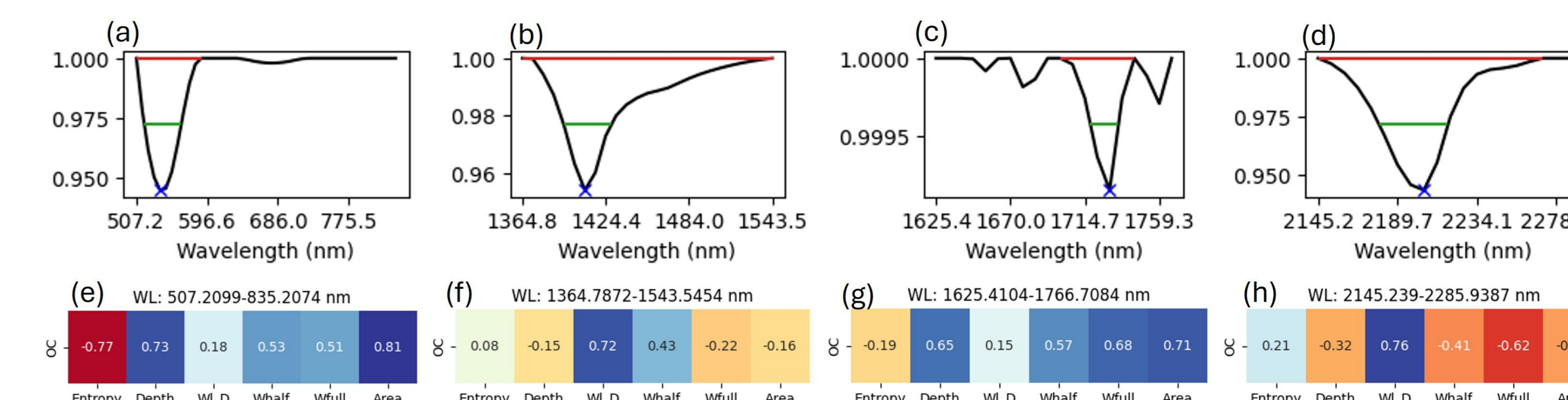
Mutual information estimated between SOC and soil spectra for (a) OR, (c) CR, and (e) FD of the LAB (black) and LUCAS (blue) datasets.



Performance comparison of machine learning models (LASSO, PLSR, and RF) for predicting SOC using LAB and LUCAS spectral datasets. Box plots show the distribution of R^2 values from 100 model iterations for three spectral preprocessing methods: OR, CR, and FD.



Data stratification Analysis (a) SOC distribution across SOC-based groups. (b) Mutual information between spectral features and SOC within each group. (c) R^2 from 100 LASSO runs per group ((d) Relationship between R^2 and importance-weighted mutual information (IWMI), indicating the link between feature relevance and model performance.



Continuum Removal (CR) feature analysis. (a–d) CR spectra across regions showing absorption depth (blue x), half-width (green), and full-width (red). (e–h) Heatmaps of correlations between CR features and SOC, highlighting informative spectral regions.

Conclusion

- Spectral preprocessing enhances data comparability and feature identification of soil organic carbon
- Data distribution governs model stability and generalization in estimation of soil properties

Acknowledgements

The authors acknowledge funding support from the Indian Institute of Science and the Prime Minister's Research Fellowship, Ministry of Education, Government of India.