AI for Life Sciences

AI4LS Challenge #1: AI for Soil Health

Predicting and visualizing soil microbial biodiversity and its drivers

Background

Soils are living ecosystems that sustain plants, animals, and humans. Healthy soils are important to protect the climate, biodiversity, land integrity, and food security. Large-scale monitoring efforts are underway to better characterize the chemical and biological factors involved in land use changes and their effect on soil health. For example, it is well understood that intensively managed agricultural land stores much less carbon than natural soil. It is however still unclear how much different types of land use change affect the magnitude of soil degradation in different ecosystems across the globe, and how much the level of degradation depends on other environmental or biological factors. The wealth of public data available on land use and many different soil characteristics across Europe, including biotic characteristics such as the abundance of different species in the ecosystem, make Artificial Intelligence / Machine Learning approaches appealing to provide novel insights in this area.

Challenge

Currently, soil microbial diversity and its relationship with environmental parameters is analyzed by estimation of taxonomic profiles, quantification of alpha and beta diversity, and use of various multivariate statistical approaches such as cluster analysis, redundancy analysis, linear discriminant analysis and generalized linear models (see here for examples: https://sites.google.com/site/mb3gustame).

In this challenge, participants are asked to go beyond these techniques and to apply Artificial Intelligence / Machine Learning methods to develop a deeper understanding of the role of the soil microbiome to contribute to soil health. You should use the Land Use / Cover Area frame statistical Survey Soil (LUCAS Soil) dataset which is a European Union effort to survey European topsoil characteristics and Microbiome sequence data from the LUCAS Soil Survey for prokaryotes, fungi, and microeukaryotes.

The datasets mentioned above are available for download in the Main Datasets and Resources section of this document. Any other datasets not mentioned here may also be used to improve your analyses and models.

The challenge has two major goals: (1) Prediction of soil biodiversity and exploration of its environmental drivers under different land-use systems, and (2) visualization of soil biodiversity and these drivers. In addition to addressing specific topics outlined below, participants should develop an interactive interface for visualization of microbial diversity and other features in the data.

The visualization should include geographical distributions of features and predictions as well as visualizations that provide informative summaries of the data.

Specific topics that may be addressed include:

- Can you develop models to differentiate unmanaged and managed lands, as well as environmentally impacted soils from non-impacted soils?
- Can you identify ways of organizing/clustering microbial species to capture variation in soil health indicators?
- Can you identify physical, chemical, and biological features that are predictors of soil conditions?
- Can you determine the relationships between physical, chemical, and biological features in order to:
    - Generate hypotheses about co-linear features?
    - Determine main parameters or predictors of soil health?
    - Make robust and simplified models?

- Can you detect phylogenetic signals or species traits of microbes indicative of soil health status or other ecosystem characteristics?
- Can data from other microbiomes (gut, marine, wastewaters, etc…) contribute to our understanding of soil microbiomes?
- Does incorporating microbial diversity improve predictions of soil carbon storage?
- Can you extrapolate/infer missing data in microbiome data to increase the resolution of the spatial variation of the microbiome?

Presentation of Results and Assessment

Teams should prepare a 10 min video describing their approach, including the datasets and methods used. You should summarize your main results and interpretations and propose the next possible scientific or analytical/technological steps arising from your work.

Teams will be evaluated based on the quality of their analysis, conceptual understanding, technological implementation, and communication skills.

Main Datasets and Resources

- Land Use / Cover Area frame statistical Survey Soil (LUCAS Soil) datasets are available for download here.
- Microbiome sequence data from the LUCAS Soil Survey for prokaryotes, fungi, and microeukaryotes will be available shortly, you will be notified via email and Discord.

Other Datasets and Resources

- National Science Foundation's National Ecological Observatory Network (NEON): https://data.neonscience.org/data-products/explore (US)
- Fungi Database: https://fungidb.org/fungidb/app
- Global Biodiversity Information Facility: https://www.gbif.org/
- Soil structure library: https://structurelib.ufz.de/lit/

- European fluxes database: http://www.europe-fluxdata.eu/
- European Commission Statement on Soil Health: https://environment.ec.europa.eu/topics/soil-and-land/soil-health_en
- USDA Description of Concepts and Parameters to Define Soil Health: https://www.nrcs.usda.gov/conservation-basics/natural-resource-concerns/soils/soil-health/soil-health-assessment

Relevant Literature

Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., ... & Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, *349*(6251), 970-973.

Köninger, J., Ballabio, C., Panagos, P., Jones, A., Schmid, M. W., Orgiazzi, A., & Briones, M. J. (2023). Ecosystem type drives soil eukaryotic diversity and composition in Europe. *Global Change Biology*, *29*(19), 5706-5719.

Labouyrie, M., Ballabio, C., Romero, F., Panagos, P., Jones, A., Schmid, M. W., ... & Orgiazzi, A. (2023). Patterns in soil microbial diversity across Europe. *Nature Communications*, *14*(1), 3311.

Lehmann, J., Bossio, D. A., Kögel-Knabner, I., & Rillig, M. C. (2020). The concept and future prospects of soil health. *Nature Reviews Earth & Environment*, *1*(10), 544-553.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science*, *69*(1), 140-153.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... & Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*(6213), 1256688.