



## **Data-mining analysis of the global distribution of soil carbon in observational databases and Earth system models**

Shoji Hashimoto (1,2), Kazuki Nanko (1), Boris Āupek (3), and Aleksi Lehtonen (3)

(1) Forestry and Forest Products Research Institute, Tsukuba, Japan (shojih@ffpri.affrc.go.jp), (2) The University of Tokyo, Bunkyo, Japan, (3) Natural Resources Institute Finland, Helsinki, Finland

Soil is the largest organic carbon pool in terrestrial ecosystems, and future climate change will dramatically change the soil organic carbon (SOC) dynamics. Moreover, the change is considered to affect the future climate. Observational global SOC databases are often used as benchmarks to examine the performance of Earth system models (ESMs); however, recent studies reported that SOC distributions simulated by ESMs and those of observational databases did not agree well. At present, the specific key processes/factors causing the mismatch are unknown. In this study, we applied a data-mining analysis (Boosted Regression Trees– BRT) to both SOC in observational databases and SOC outputs from 15 ESMs (Coupled Model Intercomparison Project; CMIP5), and sought to identify the influential factors that govern global SOC distributions and causes of mismatch between observational databases and ESMs' outputs. We examined the effects of 13 variables/factors categorized into five groups (climate, soil property, topography, vegetation, and land-use history). At a global scale, a comparison of the influential factors from ESMs and observational databases revealed that the most distinct differences between the SOC from the observational databases and ESMs were the low clay content and CN ratio contributions, and the high NPP contribution in the ESMs. The results of this study will improve the modeling of terrestrial carbon dynamics in ESMs. Process oriented models and their outputs are highly complicated and often difficult to quantitatively compare with benchmarks in detail; this study also reveals how a data-mining algorithm can be used to assess model outputs.

Hashimoto, S., K. Nanko, B. Āupek, A. Lehtonen (2017) Data-mining analysis of the global distribution of soil carbon in observational databases and Earth system models. *Geoscientific Model Development*, 10, 1321-1337. <https://doi.org/10.5194/gmd-10-1321-2017>