PROcess-guided deep learning and DAta-driven modelling (PRODA) uncovers key mechanisms underlying global soil carbon storage

Feng Tao¹, Yuanyuan Huang², Bruce A. Hungate³, Xingjie Lu⁴, Toby D. Hocking³, Umakant Mishra⁵, Gustaf Hugelius⁶, Xiaomeng Huang¹, and Yiqi Luo²

¹Tsinghua University
²CSIRO Oceans and Atmosphere
³Northern Arizona University
⁴Sun Yat-sen University
⁵Sandia National Laboratories
⁶Stockholm University

Soil carbon storage is a vital ecosystem service. Yet mechanisms that regulate global soil organic carbon (SOC) dynamics remain elusive. Here we explicitly retrieve the spatial patterns of key biogeochemical mechanisms and their regulation pathways on SOC storage using the novel PROcess-guided deep learning and Data-driven modelling (PRODA) approach. PRODA integrates data assimilation, deep learning, big data with 54,073 globally distributed vertical SOC profiles, and the Community Land Model version 5 (CLM5) to best represent and understand global soil carbon cycle. The PRODA-optimised CLM5 can represent 56±2% spatial variation of SOC across the world. Among all the mechanisms we explored in this study, microbial carbon use efficiency (CUE) emerges as the most critical regulator of global SOC storage. Increasing CUE, where more carbon flow is channelled into stabilisation, coincides with decreasing temperature and favours SOC accrual. Global sensitivity analysis further confirms the CUE, surpassing carbon input and decomposition, as the primary driver to SOC storage and its spatial variation. An increase of CUE by 1% from its standing value will lead to an additional 76±3 petagrams global SOC accumulation. We conclude that how efficiently soil microorganisms utilise organic carbon in metabolism is central to global SOC stabilisation. Understanding detailed processes underlying CUE and its environmental dependence will be critical in accurately describing soil carbon dynamics and its feedbacks to climate change.