Testing machine-learning methods toward improvement of terrestrial CO2 flux estimations

*Zhiyan Liu[1], Kazuhito Ichii[1]

1. Center for Environmental Remote Sensing Chiba University, Chiba University

Data-driven approach is effective for upscaling observation network data of terrestrial carbon fluxes. In this study, we estimated terrestrial gross primary productivity (GPP) and net ecosystem exchange (NEE) across the globe using machine learning methods –random forest regression (RF) with MODIS collection 6 datasets after previous studies used collection 5 [e.g. Ichii et al. 2017; Tramontana et al. 2016; Kondo et al. 2015]. Furthermore, we introduced lag effects of input parameters (simply input as a new column) on GPP and NEE in RF models and this lag effect include either historical or future (lag within a month) remote sensing based input parameters. Because the ecosystem is a dynamic system, which current status is affected by historical factors and meanwhile will be observed in the future as some other physics or biochemical quantities. Thus, both processes may include the lag effect which mechanism are also different. We also attempt to adopt a deep learning method called RNN (Recurrent Neural Network and its sub-types) in which data flows by each time step and make some comparision. Site-level RF experiments showed that the lagged parameter improved the accuracy of NEE estimation (R^2 0~0.06 increasing for total Koppen climate types). The lagged parameter worked well in site-level anomalies, particularly, lag effect by shortwave solar radiation or land surface temperature in the night. The estimated annual anomaly variations in GPP and NEE showed a good consistency with independent model-based estimations (TRENDY V6). In the future research, the lag-effected will be a more important factor for ecosystem’ s prediction and estimation.

Keywords: Terrestrial GPP and NEE, MODIS Collection 6, Time Lag/Memory Effect, Data Driven