Predicting carbon dioxide and energy fluxes with empirical approaches in FLUXNET.

Gianluca Tramontana (1), Martin Jung (2), Christopher R. Schwalm (3), Kazuhito Ichii (4,13), Gustau Camps-Valls (5), Botond Ráduly (6,1), Markus Reichstein (2), M. Altaf Arain (7), Alessandro Cescatti (8), Gerard Kiely (9), Lutz Merbold (10,14), Penelope Serrano-Ortiz (11), Sven Sickert (12), Sebastian Wolf (10), and Dario Papale (1)

(1) Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), Univeristy of Tuscia, Viterbo, Italy, (2) Max Planck Institute for Biogeochemistry, Jena, Germany, (3) Woods Hole Research Center, Falmouth MA, USA, (4) Department of Environmental Geochemical Cycle Research, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, (5) Image Processing Laboratory (IPL), Universitat de València, Paterna (València), Spain, (6) Department of Bioengineering, Sapientia Hungarian University of Transylvania, Miercurea Ciuc, Romania, (7) School of Geography and Earth Sciences, McMaster University, Hamilton (Ontario), Canada, (8) European Commission Joint Research Centre, Directorate for Sustainable Resources, Ispra,Italy, (9) Civil & Environmental Engineering and Environmental Research Institute, University College, Cork, Ireland, (10) Department of Environmental Systems Science, Institute of Agricultural Sciences, Zurich, Switzerland, (11) Department of Ecology, University of Granada, Granada, Spain, (12) Computer Vision Group, Friedrich Schiller University Jena, Jena, Germany, (13) Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan, (14) Mazingira Centre, Livestock Systems and Environment, International Livestock Research Institute (ILRI), Nairobi, Kenya

Global spatio-temporal fields of land-atmosphere fluxes derived from data-driven models and eddy covariance measurements can complement simulations by process-based Land Surface Models. Furthermore, they are also increasingly used for analyzing variations of the global carbon and energy cycles. However, while a number of strategies for empirical models with eddy covariance flux data have been applied, a systematic intercomparison of these methods is missing so far. Here, we report the results of a cross-validation experiment for predicting carbon dioxide, latent heat, sensible heat and net radiation fluxes, across different ecosystem types. That experiment was performed in the context of the FLUXCOM activities that aims at providing an array of improved data-driven flux products. Empirical models were derived by eleven machine learning (ML) methods from four different classes (kernel methods, neural networks, tree methods, and regression splines). Fluxes data were taken by more than 200 eddy covariance study sites over the globe. Two complementary experimental setups have been carried out: (1) 8-day average fluxes based on remotely sensed data, and (2) daily mean fluxes based on meteorological data and mean seasonal cycle of remotely sensed variables. The pattern of predictions from different ML and experimental setups were highly consistent. Instead there were systematic differences in performance among the fluxes, with the following ascending order: net ecosystem exchange ($R^2 < 0.5$), ecosystem respiration ($R^2 > 0.6$), gross primary production ($R^2 > 0.7$), latent heat ($R^2 > 0.7$), sensible heat ($R^2 > 0.7$), net radiation ($R^2 > 0.8$). The ML methods predicted very well the across site variability and the mean seasonal cycle of the observed fluxes ($R^2 > 0.7$), while the 8-day deviations from the mean seasonal cycle were not well predicted ($R^2 < 0.5$). Fluxes were better predicted at forested sites (excepting the evergreen broadleaved forest) and in the temperate or boreal climate sites than at ones in extreme climates or less represented by training data (e.g. the tropics). The evaluated large ensemble of ML based empirical models were used to derive two complementary sets of products (under evaluation) having enhanced spatial and temporal resolution: a 5 min spatially and 8 day temporally resolved product driven solely by remote sensing based variables, and a daily (vegetation type specific) product at 0.5° driven by meteorological data and mean seasonal cycle remote sensing based variables.