Geophysical Research Abstracts
Vol. 21, EGU2019-10731-1, 2019
EGU General Assembly 2019

CH4 flux gap-filling approaches for eddy covariance data: a comparison of three machine learning algorithms and marginal distribution sampling method with and without principal component analysis

Yeonuk Kim (1), Mark Johnson (1,2), Sara Knox (3), Andrew Black (4), Higo Dalmagro (5), Minseok Kang (6), Joon Kim (6,7,8), Youngryel Ryu (7,8,9), and Dennis Baldocchi (10)

(1) Institute for Resources Environment and Sustainability, University of British Columbia, Vancouver, Canada, (2) Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, Canada, (3) Department of Geography, University of British Columbia, Vancouver, Canada, (4) Faculty of Land and Food Systems, University of British Columbia, Vancouver, Canada, (5) Environmental Sciences Graduate Program, University of Cuiabá, Cuiabá, Brazil, (6) National Center for AgroMeteorology, Seoul, South Korea, (7) Department of Landscape Architecture & Rural Systems Engineering, Seoul National University, Seoul, South Korea, (8) Interdisciplinary Program in Agricultural & Forest Meteorology, Seoul National University, Seoul, South Korea, (9) Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, (10) Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, United States

Within the past decade, flux towers which incorporate methane flux (FCH4) measurements have increased following the development of a robust and low-power CH4 gas analyzer. As is the case with other trace gases, CH4 flux data based on the eddy covariance technique commonly includes data gaps due to unsuitable atmospheric conditions and system failures, with FCH4 often having more gaps than CO_2 and H_2O fluxes. However, there is still no consensus on FCH4 gap-filling method. Some researchers have applied a marginal distribution sampling (MDS) algorithm for FCH4 gap-filling, which is a standard gap-filling method for CO_2 and water fluxes. Others have tested and applied machine learning (ML) algorithms like artificial neural networks to gap-fill FCH4 data.

One of the major challenges in FCH4 gap-filling using both MDS and ML algorithms concerns selection of input variables. Since MDS has been widely used in the Fluxnet community by virtue of a robust online algorithm and an R package, standardizing the selection of MDS input variables for gap-filling FCH4 is important. Nevertheless, discussions in the literature on MDS input parameters are rare. Likewise, use of ML algorithms for FCH4 gap-filling has increased with limited discussions on input selection. At present, we are also not aware of a comprehensive comparison between commonly used ML algorithms such as artificial neural networks (ANN), support vector machine (SVM), and random forest (RF) for FCH4 gap-filling.

In this study we compare performance of MDS and the three ML algorithms (ANN, SVM, RF) using different combinations of ancillary parameters. In addition, we applied principal component analysis (PCA) as an input to the algorithms to resolve multi-parameter dependency of FCH4. We applied this approach to 5 benchmark FCH4 datasets from temperate and tropical wetlands and rice paddies. For performance evaluation, various artificial gap scenarios were added to FCH4 data and then gap-filled using MDS and ML algorithms, with and without PCA. Performance of the algorithms was measured using Taylor diagrams, and residual analysis was also conducted to understand limitations of each algorithm and input decision. Results indicate that PCA improves the performance of MDS compared to traditional MDS inputs (i.e. global radiation, air temperature, and vapour pressure deficit) and alternative MDS inputs (i.e. soil temperature, water table height, and net ecosystem exchange). On the other hands, all of the ML algorithms show better performance without PCA when all available biophysical variable are selected as inputs. As for the comparison among ML algorithms, RF was found to outperform other techniques for the 5 benchmark sites and all gap scenarios. In this presentation, we also discuss how annual CH4 budgets vary in relation to gap-filling methods based on biases introduced through the various gap-filing approaches.