World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:14, No:06, 2020

Ensemble Machine Learning Approach for Estimating Missing Data from CO_2 Time Series

Authors : Atbin Mahabbati, Jason Beringer, Matthias Leopold

Abstract : To address the global challenges of climate and environmental changes, there is a need for quantifying and reducing uncertainties in environmental data, including observations of carbon, water, and energy. Global eddy covariance flux tower networks (FLUXNET), and their regional counterparts (i.e., OzFlux, AmeriFlux, China Flux, etc.) were established in the late 1990s and early 2000s to address the demand. Despite the capability of eddy covariance in validating process modelling analyses, field surveys and remote sensing assessments, there are some serious concerns regarding the challenges associated with the technique, e.g. data gaps and uncertainties. To address these concerns, this research has developed an ensemble model to fill the data gaps of CO_2 flux to avoid the limitations of using a single algorithm, and therefore, provide less error and decline the uncertainties associated with the gap-filling process. In this study, the data of five towers in the OzFlux Network (Alice Springs Mulga, Calperum, Gingin, Howard Springs and Tumbarumba) during 2013 were used to develop an ensemble machine learning model, using five feedforward neural networks (FFNN) with different structures combined with an eXtreme Gradient Boosting (XGB) algorithm. The former methods, FFNN, provided the primary estimations in the first layer, while the later, XGB, used the outputs of the first layer as its input to provide the final estimations of CO_2 flux. The introduced model showed slight superiority over each single FFNN and the XGB, while each of these two methods was used individually, overall RMSE: 2.64, 2.91, and 3.54 g C m^{-2} yr^{-1} respectively (3.54 provided by the best FFNN). The most significant improvement happened to the estimation of the extreme diurnal values (during midday and sunrise), as well as nocturnal estimations, which is generally considered as one of the most challenging parts of CO_2 flux gap-filling. The towers, as well as seasonality, showed different levels of sensitivity to improvements provided by the ensemble model. For instance, Tumbarumba showed more sensitivity compared to Calperum, where the differences between the Ensemble model on the one hand and the FFNNs and XGB, on the other hand, were the least of all 5 sites. Besides, the performance difference between the ensemble model and its components individually were more significant during the warm season (Jan, Feb, Mar, Oct, Nov, and Dec) compared to the cold season (Apr, May, Jun, Jul, Aug, and Sep) due to the higher amount of photosynthesis of plants, which led to a larger range of CO_2 exchange. In conclusion, the introduced ensemble model slightly improved the accuracy of CO_2 flux gap-filling and robustness of the model. Therefore, using ensemble machine learning models is potentially capable of improving data estimation and regression outcome when it seems to be no more room for improvement while using a single algorithm.

Keywords : carbon flux, Eddy covariance, extreme gradient boosting, gap-filling comparison, hybrid model, OzFlux network

Conference Title : ICMLMEES 2020 : International Conference on Machine Learning Methods in Ecological and Environmental Sciences

Conference Location : Copenhagen, Denmark

Conference Dates : June 11-12, 2020

Open Science Index, Environmental and Ecological Engineering Vol:14, No:06, 2020 publications.waset.org/abstracts/122496.pdf