A hybrid approach to mountain forest management using Earth observing system and machine learning

Polina Sysoeva, Peter Khaiter

School of Information Technology, York University, Toronto, Canada
Email: pkhaiter@yorku.ca

Abstract: Carbon sequestration is one of the key ecosystem services provided by forests. Its accurate assessment is an important task as the amount of carbon a forest can sequestrate is an indicator of ecosystem health as well as a contribution to maintaining of environmental sustainability. Nowadays, remote sensing has become a widely applied technique to acquire data for ecosystem service assessment (ESA). However, in part due to its abundancy, it is impossible to get insights from the data without the use of advanced computational algorithms, such as machine learning (ML).

A recent review analyzed the history of Light-Use Efficiency models that calculate the capacity of the ecosystem to store carbon measured by the gross primary productivity (GPP) indicator. Among other issues, important implications of choosing the correct data sources have been discussed, such as FLUX towers, remotely sensed data or in situ observations as well as proper preprocessing for the confidence of the policy- and decision-makers in the model results. The novelty of our approach lies in using only satellite data, thus avoiding flux towers and in situ measurements, and stepping away from any one specific vegetation proxy but relying on multiple indexes instead. This serves as a means of correcting possible model limitations and allows to generalize the approach to territories where the possibility of in situ measurements is either significantly limited or completely impossible due to their remoteness or inaccessibility, as the case is with mountain forests.

The methodology proposed includes collecting MODIS data from Google Earth Data Catalogue, resampling it and using time series interpolation to impute missing observations. We explore the properties of data and its suitability to make predictions by means of inferential statistics and time series analysis. Three sets of experiments are conducted incorporating ensemble learning methods, and transformers as well as studying explainability of the results via SHAP values. Predictions from the models for two regions of mountain forests in the Kashmir Valley, India are adopted for ESA through the market values of carbon obtained from different Emission Trading Systems (ETS).

Random Forest Regression has shown the best performance in forecasting GPP but predictions from XGBoost Regression are better explainable due to the algorithm's immunity to multicollinearity in data training. SHAP value analysis revealed that the Enhanced Vegetation Index (EVI) is a predictor that affected the GPP the most. Our model has demonstrated that two flood events that occurred in 2015 resulted in carbon sequestration reduction in the Dachigam National Park amounting to a sunk cost of $1.3 million in EU ETS prices. Similarly, it showed that the predicted alternative costs of the mining activities in the village of Khrew in 2023 are equivalent to $1.2 million in EU ETS prices.

The approach we propose contributes to the field of sustainable mountain forest management and will assist policy- and decision-makers in ecosystem service assessment by integrating the predictions with economic estimates and indicators from the area of the emissions trading.

Keywords: *Ecosystem service assessment, Earth observing system, geographic information systems, machine learning*