26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Evaluation of empirical and radiative transfer modelling for estimating biochemistry using remote sensing data

A Simic Milas [a] **and MMT Alam** [a]

[a] *School of Earth, Environment and Society, Bowling Green State University, 190 Overman Hall, Bowling Green, OH, USA*
Email: asimic@bgsu.edu

Abstract: This study evaluates two key approaches for estimating canopy chlorophyll content (CCC) and canopy nitrogen content (CNC), both essential indicators for precision agriculture: (1) empirical machine learning regression algorithms (MLRAs), and (2) hybrid methods that integrate physically based radiative transfer models (RTMs) with MLRAs. The objective is to determine which method more effectively supports crop monitoring using remote sensing data. The research was conducted at the Kellogg Biological Station (KBS) in Michigan, USA, a long-term agricultural research site with extensive ecological and crop-related datasets. Three modeling scenarios were assessed: (1) multispectral satellite imagery (Landsat 7, RapidEye, and PlanetScope); (2) UAV–satellite data fusion; and (3) hyperspectral modeling using data from the EnMAP satellite. In the empirical modeling framework, six different MLRAs were tested. Additionally, non-vegetated spectra (pixels from bare soil, roads, water bodies) were added to the training dataset to enhance the accuracy of the MLRAs. The 3-fold cross-validation method was employed to determine their accuracy. The hybrid modeling approach—combining RTM (specifically PROSAIL) with MLRAs—was implemented using ARTMO (Automated Radiative Transfer Models Operator), a software tool that facilitates the automated training and evaluation of both empirical and hybrid models.

Kernel Ridge Regression (KRR) and Gaussian Process Regression (GPR) consistently delivered strong performance, especially when integrated with UAV-derived structural metrics such as vegetation index, canopy height, and leaf area index. CCC estimates achieved an R^2 of 0.89 and an RMSE of approximately 9 µg/cm² when paired with RapidEye data. With hyperspectral EnMAP data, KRR achieved even higher accuracy, reaching R^2 values above 0.93 for both CCC and CNC.

MLRAs like KRR and GPR have proven effective in capturing complex, non-linear relationships between spectral data and biochemical variables, particularly when enhanced with structural and spectral inputs from high-resolution sensors, including UAVs and hyperspectral satellites. RTMs, such as PROSAIL—which combines the PROSPECT leaf model with the SAIL canopy model—simulate plant reflectance based on physical properties, though they require computationally intensive simulations and extensive parameterization. In summary, empirical MLRA models demonstrated high accuracy and scalability for estimating crop biochemical traits. Hybrid models offer added value in heterogeneous or data-scarce environments due to their enhanced generalization capability. While empirical methods are well-suited for small and consistently managed agricultural areas, hybrid approaches are better equipped to handle variability and limited field data. Both methods contribute significantly to advancing remote sensing applications in sustainable agriculture.

Keywords: *Hybrid model, RTM, machine learning, crops, biochemistry*