A COMPARISON OF SPECTRAL INDICES AND MACHINE-LEARNING METHODS FOR THE ESTIMATION OF LEAF CHLOROPHYLL CONTENT

Bachelor's thesis

Performed by: XIE Chu, student of the 4th year, group GK-42
Scientific supervisor: PhD, assoc. prof. Olena AHAPOVA
V.N. Karazin Kharkiv National University, 2024
Specialty: 106 Geography
Educational program: Cartography, Geinformatics and Cadastre

ABSTRACT

The relevance of the research. Accurately and efficiently retrieving leaf chlorophyll content (LCC) is a crucial technique for achieving intelligent crop breeding and precision agriculture. However, traditional methods for measuring leaf chlorophyll content require substantial time and manpower. Remote sensing inversion techniques provide a rapid, non-destructive, and relatively straightforward approach for obtaining leaf chlorophyll content.

The object of the research is vegetation and the leaf chlorophyll content, and **the subject** is machine-learning method with spectral indices for the estimation of leaf chlorophyll content.

The purpose of the research is to compare spectral indices and machine-learning methods for the estimation of leaf chlorophyll content and determine the optimal ones.

To achieve the purpose of the work, the following **tasks** were set:

1. Construction of PROSPECT simulation dataset and measured dataset: On one hand, representative measured datasets from different ecosystems are collected. On the other hand, based on the PROSPECT-5 model, ideal leaf reflectance in the range of 400 to 2500 nm is simulated under different biochemical and structural parameters, and sensitivity analysis is conducted on the PROSPECT simulation dataset.

2. Division of datasets based on LCC content: The datasets are arranged in order of increasing chlorophyll content, and then divided into training set (60%) and validation set (40%) using uniform sampling method, ensuring coverage of data types from different ecosystems.

3. Construction of LCC inversion model based on vegetation indices and accuracy validation: A univariate linear regression model based on vegetation indices is constructed, and the inversion results of the vegetation index model are compared with the measured chlorophyll content to obtain the optimal spectral index for LCC inversion.

4. Evaluation of machine learning model inversion accuracy on datasets: The k-fold cross-validation method is adopted to train and optimize the hyperparameters of machine learning methods using the measured dataset in the training set. Subsequently, the trained model predicts LCC in the validation set and compares it with the measured LCC in the validation set to determine the best LCC inversion machine learning method.

5. Comparison of advantages and disadvantages of LCC remote sensing inversion methods based on spectral indices and machine learning: Evaluation of the accuracy and applicability of spectral indices and machine learning methods in leaf chlorophyll content inversion, and determination of the optimal LCC remote sensing inversion method under different datasets.

Research methods. In this study, representative datasets covering various ecosystems were utilized to compare the accuracy of multiple published spectral indices and machine learning methods for retrieving vegetation leaf chlorophyll content. Additionally, sensitivity analyses were conducted using the PROSPECT radiative transfer model. The results showed that the modified difference ratio index (MDRI) exhibited the strongest linear relationship with measured leaf chlorophyll content (Adjusted R2 = 0.8262; $p < 0.001$), outperforming other red-edge-based indices and machine learning methods. Meanwhile, red-edge indices such as CIred-edge demonstrated even stronger linear relationships in simulated datasets (r2=0.7788; Adjusted R2 = 0.9247; $p < 0.001$). Moreover, in terms of machine learning methods, Random Forest (RF) and Extratree regression models showed strong predictive capabilities in a wide range of measured datasets ($R2 = 0.7616$; nRMSE=10.25μg/cm2; MAE=6.37μg/cm2), while the Extratree regression model performed exceptionally well in simulated datasets ($R2 = 0.99$; nRMSE= 0.3493μg/cm2; MAE= 0.0601μg/cm2). This study provides a comprehensive assessment of the application of spectral indices and machine learning in leaf chlorophyll content retrieval, highlighting the advantages of machine learning methods in mining hyperspectral information and contributing to a deeper understanding of linear and nonlinear models for vegetation biochemical parameter inversion.

Structure of the work. The qualification work consists of introduction, four chapters and conclusions. The reference list includes 55 positions. The thesis is laid out on 47 pages. Contains 10 figures and 9 tables.