

Peer review report on “Assessing forest properties with data-driven vegetation indices: insights from 900,000 forest stands”

Comments to Authors:

Overview and Major Comments:

This manuscript presents a comprehensive and methodologically innovative study that systematically explores the potential of data-driven vegetation indices (VIs) to estimate key forest properties (biomass, LAI, GPP, and NPP). By coupling an individual-based forest dynamics model (FORMIND) with a multilayer radiative transfer model (mSCOPE) and applying a Monte Carlo sampling strategy, the authors generate an exceptionally large and well-structured synthetic dataset. The analysis of all possible two-wavelength combinations across the 400–2400 nm range, combined with an explicit treatment of multiple uncertainty sources, represents a substantial advance over existing studies.

Overall, the manuscript is scientifically sound, clearly written, and highly relevant to the remote sensing and forest ecology communities, particularly in the context of emerging hyperspectral satellite missions such as EnMAP. The study offers valuable conceptual insights into wavelength selection, index design, and uncertainty robustness. I think the manuscript suitable for publication after some revisions, mainly aimed at clarifying applicability and ensuring reproducibility.

Major comments

1. The introduction could be strengthened by improving accessibility for a broad audience. A greater emphasis on the ecological motivation would be beneficial. The authors could clarify why forest parameters such as forest biomass, LAI, GPP, and NPP are critical variables and why their large-scale estimation remains challenging.
2. While the comprehensive coupling of forest model and radiative transfer model is impressive, the manuscript would be more convincing if the authors could demonstrate that FORMIND reasonably represents real forest conditions. For example, how do the observation-based estimates of forest properties at sites such as “Hohes Holze” (i.e., biomass, LAI, GPP, NPP) compare with the simulated ranges generated by FORMIND? Are all observed values captured within the model’s simulated distributions?

Minor comments:

Line 5: “two wavelengths (400 nm-2400 nm)” → “two wavelength (within 400 nm -2400 nm), to avoid the confusion.

Line 116: You may specify the meaning of the ODM and the correct unit for different variables.

Lines 117-118: Consider briefly explaining the interpretation of DBH entropy values—for example, does a more negative value indicate lower heterogeneity?

For Table 1: It would be helpful to add references or explanations regarding the rationale for selecting the wavelengths used in classical indices.

Lines 179-180: what are the criteria to select the “separating thresholds” for biomass and DBH entropy?

Lines 341-342: The study would be further strengthened if the developed hybrid model could be validated against sites with hyperspectral observations and observation-based estimates of forest properties in future work.