Dataset for the comparative analysis of the structural characteristics of terrestrial laser scanning, digital aerial photogrammetry, and field-based inventory

Motivation for Dataset Creation

Questions	Responses
Why was the dataset created?	The dataset was created as part of a project aimed at developing an integrated monitoring framework using sound-based detection systems, Internet of Things (IoT) technologies, and remote sensing to support the detection of illegal logging in Kenya. Within this broader project, the dataset focuses on evaluating remote sensing approaches for assessing vegetation structural characteristics required to support biomass estimation and carbon monitoring. Specifically, the dataset was designed to compare terrestrial laser scanning (TLS) and drone-based digital aerial photogrammetry (DAP) in their ability to capture three-dimensional vegetation structure. TLS provides high-resolution information on internal canopy structure, whereas DAP enables spatially extensive characterization of the upper canopy. To improve feasibility for large-scale or operational monitoring, structural metrics were derived using simplified acquisition and processing strategies, including a single-scan TLS approach and a digital terrain model–independent normalization approach for DAP were tested. These approaches reduce time, cost, and technical complexity at the expense of some measurement precision. The resulting dataset enables direct comparison of structural metrics derived from TLS and DAP and was created to assess the extent to which simplified, cost-effective remote sensing methods can capture comparable dimensions of forest structure relevant to biomass mapping and carbon monitoring.
What (other) tasks could the dataset be used for?	The dataset can be used for a wide range of applications related to the use of remote sensing in ecological and forest studies, particularly those involving terrestrial laser scanning (TLS) and drone-based digital aerial photogrammetry (DAP). It is suitable for both methodological and applied research, including evaluations of data acquisition strategies, processing workflows, and the performance of different remote sensing techniques in capturing forest structural attributes. Beyond technical assessments, the dataset can support ecological analyses that rely on forest structural metrics, such as studies of aboveground biomass, carbon stocks, canopy structure, and habitat-related variables. The accompanying field inventory data further enable the calibration, validation, or comparison of remote sensing–derived estimates of forest parameters. The dataset may also be used for algorithm development, sensitivity analyses, and benchmarking simplified or cost-effective remote sensing approaches against higher-resolution structural measurements. While the dataset can provide valuable insights into the capabilities of TLS and DAP, interpretations intended to inform management or policy decisions should be confined to the environmental and geographic context of the study area, as transferability beyond this scope has not been evaluated.
Has the dataset been used for any tasks already?	The dataset has been used in a published research study by the data providers. Research article can be found here: **https://doi.org/10.1002/rse2.70047**
Who created the dataset?	The published dataset was compiled by Magnus Onyiriagwu under the supervision of Delphine Clara Zemp at the Conservation Biology Laboratory, University of Neuchâtel, Switzerland.
Has the dataset been used for any tasks already?	The data was used in the study by Onyiriagwu et al., (2025). Dataset and scripts are published in Zenodo: https://doi.org/10.5281/zenodo.17038271

Who funded the creation of the dataset?	The project leading to the data compilation was funded by the Climate Change AI Innovation Grants Program with the support of the Quadrature Climate Foundation, DeepMind, and the Canada Hub of Future Earth (IG-2023-113).
Any other comments?	None

Dataset Composition

What do the instances that comprise the dataset represent?	Interpreting instances as variables in this case, the dataset comprises numeric values for a set of structural variables extracted from the TLS, DAP, and field inventory in forests and agroforestry landscapes in the study region. TLS and DAP metrics were derived from their respective point cloud data, while field metrics were directly measured from individual trees in the field using the traditional forest inventory techniques Each derived variable falls into one of the four vegetation structural aspects, including the vertical structure, horizontal structure, vegetation density, and structural heterogeneity.
How many instances of each type are there?	The dataset contains 6 TLS, 12 DAP, and 8 field structural metrics for each of the 36 measured plots (18 forest and 18 agroforestry plots). Each of the variables are numeric continuous variables in either metric or composite scale for the directly measured variables and the derived indices, respectively. The structural metrics from the DAP and TLS are processed outputs from their point cloud data
Is everything included, or does the data rely on external resources?	The dataset includes all the information required to reproduce results. In generating the raw data: • TLS relies on FARO scene software and the SSCI algorithm developed by Ehbrecht et al., (2017) and can be assessed on GitHub: https://github.com/ehbrechtetal/Stand-structural-complexity-index---SSCI • DAP relies on Pix4D for image transformation and protocol by Mohan et al., (2021) for deriving metrics. FARO Scene and Pix4D software are licensed software. The SSCI algorithm is free to use.
What experiments were initially run on this dataset?	In the published article: • Spearman correlation to assess the bivariate correlation between individual pairs of structural metrics • Procrustes correlation and principal component analysis to quantify the compatibility of each of the structural dimensions of the TLS, DAP, and field data with each other. • Random forest classification to quantify the performance of the TLS, DAP, and Field metrics as well as their metrics in delineating the structural characteristics of the forests and agroforestry landscapes.

Data Collection Process

How was the data collected?	Data for TLS, DAP, and field inventory were collected from 34, 30 m wide circular plots set up equally within the forest and agroforestry sites. TLS data was acquired using the terrestrial laser scanning device mounted on a tripod at a height of 1.3 m above ground. Raw TLS scans were first

	transformed into XYZ format in the FARO scene software. Then, the TLS metrics were generated by running the scans through the SSCI algorithm Raw aerial images were collected by flying an Ebee X fixed-wing drone equipped with a 3D sensor camera through the plots. Images were transformed to point cloud in the Pix4D software. Structural metrics were computed from the point cloud following the protocol detailed in Mohan et al., (2021). DAP point clouds were generated by reconstructing 2D aerial images into 3D data using the structure-from-motion technique that was implemented in the Pix4D software. Field inventory was conducted following the guidelines of the Kenyan Forestry Service
Who was involved in the data collection process?	The data was acquired through the collaborative effort of Caleb W.T. Ngaba, Anthony Macharia, Henry Muchiri, Abdalla Kisiwa, Magnus Onyiriagwu, Delphine Clara Zemp, Allan Vikiru, Zainabu Muti, with help from the Kenya Forest Research Institute (KEFRI), and the community forest association of the surrounding communities. All contributions to the idea conceptualization, methodological development, administration, analysis, drafting, data curation and funding acquisition for the project from which the dataset emanates are recognized through coauthorship in published articles. Other contributions in the data acquisition were financially enumerated and are formally acknowledged in all scientific output
Over what time-frame was the data collected?	The TLS, DAP, and Field data were simultaneously acquired between 5th and 24th January 2024
Does the dataset contain all possible instances?	Each observation in the dataset are instances of the structural attributes of the forest and agroforestry vegetation in the region acquired through a two-stage systematic sampling. The first stage comprised 6 rectangular blocks of 25 ha with equal parts of the blocks lying in the forest and agroforest landscape. In the second stage, a systematic sample of six 30 m circular plots, 3 in each of the two land-use types.
If the dataset is a sample, then what is the population?	The population is the whole extent of the Nyangores forest and the surrounding agroforestry communities.
Are there any known errors, sources of noise, or redundancies in the data	Errors in the TLS and DAP data can arise from differences in their scanning angle, sensor properties, and measurement approach. The bottom-top scanning perspective of the TLS limits its ability to capture the top of canopy attributes with great precision and can lead to underestimation of top-of-canopy metrics. In addition, a single scan TLS approach does not fully depict the 3D structure as TLS data derived by stitching multiple scans due to occlusion and data shadows. DAP cannot penetrate dense canopies, and the accuracy of its depiction of vegetation understory features declines as the canopy density increases. Hence, errors may propagate to extracted metrics for the DAP point cloud normalized using its own terrain information.
Any other comments?	

Data Preprocessing

What preprocessing/cleaning was done?	Raw aerial images were transformed to 3D point clouds using the structure-from-motion principle implemented in the Pix4D software. The 3D point clouds were normalized before data extraction using their digital terrain model (DTM).

	Raw TLS scans were transformed into XYZ format in FARO Scene software.
Was the "raw" data saved in addition to the preprocessed/cleaned data?	Both raw data and processed results are saved and stored in the local repository of the conservation biology laboratory, University of Neuchâtel.
Does this dataset collection/processing procedure achieve the motivation for creating the dataset stated in the first section of this datasheet?	The dataset fulfilled its intended purpose and can complement further studies on biomass monitoring.
Any other comments?	None

Dataset Distribution

How is the dataset distributed?	The dataset is openly available online with an assigned DOI for referencing
When will the dataset be released/first distributed?	The dataset was released on Sep 2, 2025
What license (if any) is it distributed under?	It is distributed under the Creative Commons Attribution 4.0 International
Are there any fees or access/export restrictions?	No, the data is freely available
Any other comments?	None

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset?	The data is supported, hosted, and managed by Magnus Onyiriagwu, under the supervision of Delphine Clara Zemp at the University of Neuchâtel, Switzerland.
How does one contact the owner/curator/manager of the dataset (e.g. email address, or other contact info)?	The owners can be contacted by email : Clara.zemp@unine.ch magnus.onyiriagwu@unine.ch
Will the dataset be updated? How often and by whom? How will updates/revisions be documented and communicated (e.g., mailing list, GitHub)?	The dataset will not be updated for the purpose for which it was initially conceived. Any updates to the R code scripts will be published and accessible on GitHub at https://github.com/Magchris/Compare_TLS_DAP_Field/blob/master
If others want to extend/augment/build on this dataset, is there a mechanism for them to do so? If so, is there a process for tracking/assessing the quality of those contributions?	Interested parties can contact the data owners by email. The obligation to communicate any new research aiming to use parts or all of the dataset rests on the intended user. The extent of data usage and contribution will determine whether the owners of the data should be invited for coauthorships or cited in publications.
Any other comments?	None

Legal and Ethical Considerations

If the dataset relates to people (e.g., their attributes) or was generated by people, were they	The dataset does not relate to people

informed about the data collection?	
If it relates to other ethically protected subjects, have appropriate obligations been met?	Not applicable
If it relates to people, were there any ethical review applications/reviews/approvals?	Not applicable
If it relates to people, were they told what the dataset would be used for and did they consent? What community norms exist for data collected from human communications? I	Not applicable
If it relates to people, could this dataset expose people to harm or legal action?	Not applicable
If it relates to people, does it unfairly advantage or disadvantage a particular social group?	Not applicable
If it relates to people, were they provided with privacy guarantees?	Not applicable
Does the dataset comply with the EU General Data Protection Regulation (GDPR)?	Not applicable
Does the dataset contain information that might be considered sensitive or confidential?	None
Doesthe dataset contain information that might be considered inappropriate or offensive?	None
Any other comments?	None

References	Ehbrecht, M., Schall, P., Ammer, C., & Seidel, D. (2017). Quantifying stand structural complexity and its relationship with forest management, tree species diversity and microclimate. *Agricultural and Forest Meteorology*, *242*, 1–9. https://doi.org/10.1016/j.agrformet.2017.04.012 Mohan, M., Leite, R. V., Broadbent, E. N., Jaafar, W. S. W. M., Srinivasan, S., Bajaj, S., Corte, A. P. D., Amaral, C. H. do, Gopan, G., Saad, S. N. M., Kamarulzaman, A. M. M., Prata, G. A., Llewelyn, E., Johnson, D. J., Doaemo, W., Bohlman, S., Zambrano, A. M. A., & Cardil, A. (2021). Individual tree detection using UAV-lidar and UAV-SfM data: A tutorial for beginners. *Open Geosciences*, *13*(1), 1028–1039. https://doi.org/10.1515/geo-2020-0290 Onyiriagwu, M., Leley, N., Ngaba, C. W. T., Macharia, A., Muchiri, H., Kisiwa, A., Ehbrecht, M., & Zemp, D. C. (2025). On the compatibility of single-scan terrestrial LiDAR with digital photogrammetry and field inventory metrics of vegetation structure in forest and agroforestry landscapes. *Remote Sensing in Ecology and Conservation*, *11*, 5. https://doi.org/10.1002/rse2.70047