Biodiversity Change: A Spatiotemporal Machine Learning Approach to Detect Forest Canopy Height Changes across the U.S. West Coast

Anonymous Author(s)
Affiliation
Address
email

Abstract

Monitoring forest structural changes is critical for understanding biodiversity dynamics and responding to natural disturbances such as wildfires. In this study, we present a machine learning framework that combines optical remote sensing data from HISTARFM Landsat product with spaceborne LiDAR measurements from the GEDI mission to generate forest canopy height maps and detect structural changes at 30-meter resolution across the U.S. West Coast. We developed ConvLSTM models trained using monthly composites of Landsat data and annual GEDI-derived RH98 canopy height metrics. Additionally, we explored two approaches for incorporating uncertainty information into ML modeling process. Our best-performing model achieved an RMSE of 7.536 and a Pearson's r of 0.847. Using the trained model, we generated canopy height maps for 2019 and 2020 and performed change detection by differencing these maps. Evaluation against 135 wildfire events yielded a moderate ROC AUC of 0.65. We analyzed detection errors and outlined potential improvements from both data and modeling perspectives.

1 Introduction

Biodiversity encompasses the variety of life on Earth, including the diversity of forest structures, species, and ecosystems. Forest structure, characterized by the spatial arrangement and physical attributes of vegetation, plays a critical role in shaping biodiversity patterns. Structural metrics derived from forest data are being recognized as effective proxies for biodiversity [11]. Remote sensing technologies such as Light Detection and Ranging (LiDAR) and high-resolution satellite imagery capture fine-scale structural variation over large spatial and temporal extents, enabling cost-effective biodiversity assessment in areas where field inventories are impractical. Machine Learning (ML) models, particularly Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM), are well-suited for analyzing the high-dimensional, unstructured remote sensing data [6, 14]. Together, they have been effectively applied to geospatial tasks [9]. In this study, we explored spatiotemporal ML techniques for detecting forest structural changes using remote sensing datasets from NASA missions. Our key contributions include: (1) a novel use of Hierarchical Equal Area isoLatitude Pixelation (HEALPix) for ML model validation, (2) a spatiotemporal ML model for generating wall-to-wall canopy height maps, (3) incorporation of data uncertainty into model training, and (4) a change detection analysis focused on wildfires.

2 Related Work

Traditional approaches for forest height estimation have relied heavily on airborne LiDAR, which offers precise 3D structural information but is limited by high acquisition cost and low coverage [3, 7].

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

To address these limitations, scientists and engineers have explored the use of satellite-based data, including optical imagery, Synthetic Aperture Radar (SAR), and spaceborne LiDAR. When combined with traditional ML methods, they have been used to predict forest canopy height with reasonable accuracy. However, these methods often depend on manual feature engineering and may struggle to generalize across diverse forest types and conditions [13, 5]. Deep learning models have gained attention for their ability to automatically learn hierarchical spatial and temporal features from remote sensing data. These models have been successfully applied to forest canopy height estimation tasks using both optical and LiDAR-derived inputs [1, 14]. Recent efforts have introduced advanced architectures such as UNet [10] for semantic segmentation and LSTM-based networks for modeling long-range temporal dependencies. However, model generalizability remains a key challenge, particularly when transferring across regions with varying forest compositions and characteristics. This study builds on these foundations by developing a spatiotemporal Convolutional Long Short-Term Memory (ConvLSTM) model using an optical imagery dataset derived from NASA's Landsat missions and LiDAR measurements from the GEDI mission to generate forest height maps and detect changes.

3 ML-Ready Data Sets

We focus on the Western US Pacific forests, as shown in the light-blue polygon in Figures 1 and 2, an ideal test region given their high biological and structural diversity, varied disturbance regimes, and dense coverage of training data availability. Our study targets RH98, the forest relative height at the 98^{th} percentile, retrieved at 30-meter resolution.

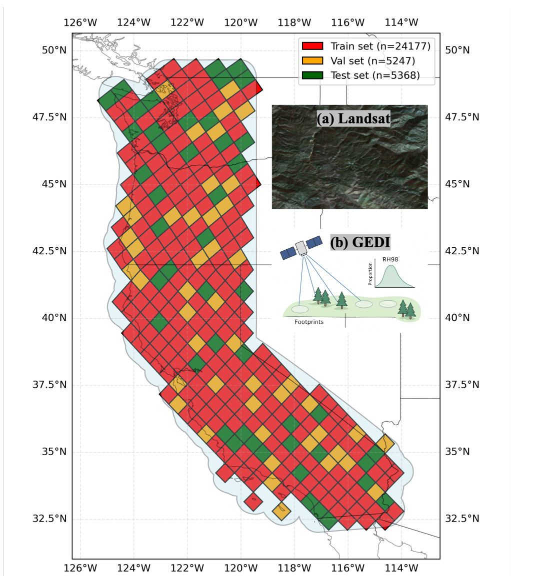

Figure 1: Train, validation, and test split

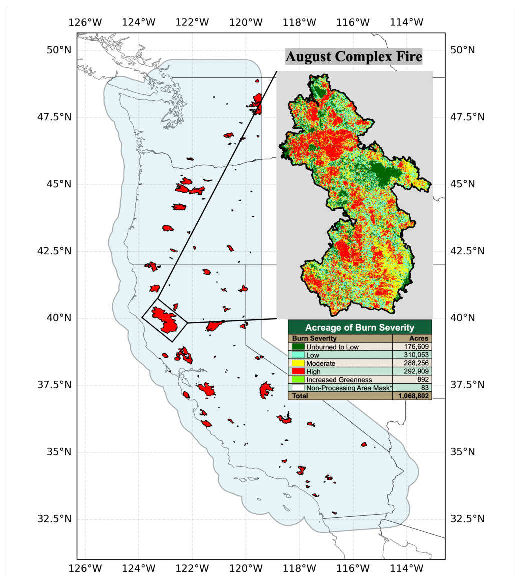

Figure 2: Wildfire events in 2020

HISTARFM Landsat Optical Imagery: The NASA Landsat program, operating since 1972, provides continuous multispectral imagery at 30-meter resolution, capturing visible, near infrared (NIR), and shortwave infrared (SWIR) bands via the Operational Land Imager (OLI) instrument. To minimize the impact of noise in raw Landsat data for ML models, we employ the HIghly Scalable Temporal Adaptive Reflectance Fusion Model (HISTARFM) Landsat dataset [15], a gap-filled monthly reflectance time series generated by fusing Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) observations. An example visualized as an RGB image is shown in Figure 1 (a). For ML modeling, HISTARFM Landsat images are tiled into arrays of size 6 x 12 x 224 x 224, where 6 corresponds to selected spectral bands (i.e., RGB, NIR, SWIR1, and SWIR2), 12 corresponds to monthly observations, and 224 x 224 corresponds to spatial dimensions.

GEDI LiDAR Measurements: The Global Ecosystem Dynamics Investigation (GEDI) mission, launched aboard the International Space Station (ISS) in 2018, provides LiDAR waveforms at ~25 meter footprints along orbital tracks between $\pm 51.6°$ latitude [2]. Measurements are sampled every 60 meters along the track, with 600 meters separating the individual laser beams, resulting in sparse observations of vegetation structure and biomass. An artistic illustration is shown in Figure 1 (b). We used the L2A Geolocated Elevation and Height Metrics product (via Google Earth Engine), filtered to vegetation growing season (April to October) to limit the snow cover effect, restricted to full-power beams with sensitivity greater than 0.9, and processed with respect to valid degradation and quality flags. The data are resampled to 30-meter resolution, aligned with the HISTARFM Landsat reference grid, and partitioned into 224 x 224 pixel tiles to use as the ground truth labels for ML models.

Wildfire Dataset: To evaluate change detection results, we compiled a wildfire dataset from the Monitoring Trends in Burn Severity (MTBS) program. The dataset includes 135 wildfire events from 2020 across California and the coastal regions of Oregon and Washington, each exceeding $1,000$ acres. Locations of the wildfire events are shown in Figure 2. Burn severity classes (e.g., see August Complex fire in Figure 2) for each event are provided as georeferenced raster files, which were merged and resampled to the HISTARFM Landsat reference grid.

4 Methodology

Spatial Validation: We used a standard train, validation, and test split methodology to evaluate the ML model. Because the tiles derived from the Landsat and GEDI data are spatially correlated, it is critical to ensure spatial independence and prevent overlap across the splits. We present a novel application of HEALPix to spatially divide the globe into equal-area regions for model validation [4]. The equal-area nature of the pixelization ensures that no regions are disproportionately represented. The pixelization schema used for this study is shown in Figure 1. The dataset was partitioned into 70% for training, 15% for validation, and 15% for testing, corresponding to $24\,177$, 5247, and 5368 tiles, respectively.

ML Training and Evaluation: We employed a supervised regression framework to estimate forest canopy height, using HISTARFM Landsat spatiotemporal tiles as input and GEDI-derived RH98 as ground truth. The backbone architecture was a ConvLSTM network, with the fully connected layers in traditional LSTMs replaced by convolution operations to learn both spatial and temporal patterns [12]. We explored three ConvLSTM variants: (1) a baseline model trained on the Landsat spectral bands (i.e., RGB, NIR, SWIR); (2) a stacked model trained with six uncertainty bands stacked onto spectral bands as input; and (3) a weighted-loss model in which the Mean Squared Error (MSE) loss is weighed by the inverse of the corresponding uncertainty values. Hyperparameters were tuned on the validation set via a small-scale grid search, with early-stopping applied to prevent overfitting. Model performance was assessed using Root Mean Squared Error (RMSE) and Pearson correlation coefficient on the train, validation, and test sets. Model bias was examined by computing the best-fitting regression line between the predicted and ground truth GEDI RH98 values.

Wall-to-wall Mapping and Change Detection: After training and evaluation, the ML model was applied to generate wall-to-wall prediction tiles across the study area. These tiles were georeferenced and mosaicked to produce a forest canopy height map at 30-meter resolution. Change detection was performed by aligning and differencing temporally distinct maps on a pixel-by-pixel basis. This simple but effective approach captures the magnitude and direction of change, with positive and negative values indicating gains and losses in forest canopy height.

5 Results and Analysis

ML Results and Analysis: The test set results of the models are summarized in Table 1. The baseline model achieved the best performance, with an RMSE of 7.536, Pearson's r of 0.847, and a regression slope of 0.80 (Figure 3). For the stacked model, stacking uncertainty bands onto the spectral bands negatively impacted the model performance. The performance degradation can be attributed to several factors. First, the inclusion of uncertainty

Table 1: Held-out test set results of the ML models.

Models	RMSE	Pearson's r	Slope
Baseline model	**7.536**	**0.847**	**0.80**
Stacked model	9.195	0.780	0.64
Weighted-loss	7.868	0.830	0.77
Weighted-loss (threshold)	7.486	0.851	0.78

bands may have introduced noisy and low-signal information that did not contribute meaningfully to model training. Unlike the spectral bands that capture physical properties, uncertainty bands represent statistical variation, which may not carry spatial or temporal patterns that are directly relevant or predictive for GEDI RH98. Second, the model may have overfitted to uncertainty bands, causing it to prioritize uncertainty-related noise over the spatial and temporal patterns in the spectral bands. Third, the model architecture may not have been optimized to interpret uncertainty bands as a distinct type of input, thus failing to leverage its intended use. The weighted-loss model performed slightly worse than the baseline on the full test set, but outperformed the baseline when evaluation was restricted to low-uncertainty samples (i.e., 90.22% test set data included). This suggests that uncertainty-weighted training improves reliability on high-quality data, though predictions in high-uncertainty regions remain degraded. All three models exhibited bias towards under-predicting tall trees. This is likely due to the imbalanced nature of the GEDI data, with tall trees severely under-represented.

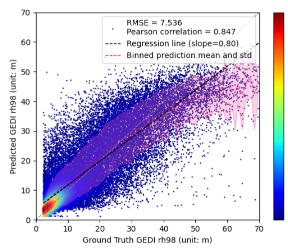

Figure 3: Test set performance for the baseline ML model

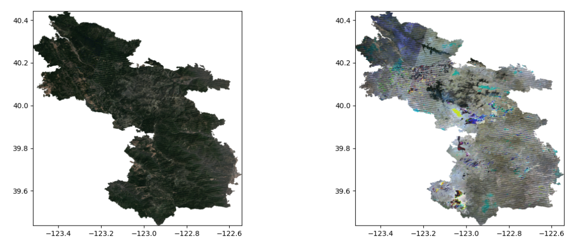

Figure 4: HISTARFM Landsat RGB images acquired in Aug. (left) and Sep. (right) 2020 for August Complex fire

Change Detection Results and Analysis To evaluate change detection, we generated wall-to-wall canopy height maps for 2019 and 2020 using the baseline ML model, and compared the differenced map with the 2020 wildfire events in Figure 2. The analysis focused on pixels labeled as the "High" and "Moderate" burn severity classes, as the "Unburned to Low" and "Low" burn classes generally do not correspond to measurable canopy height loss. The resulting Receiver Operating Characteristic (ROC) yielded an Area Under the Curve (AUC) of 0.65, which suggests that this approach has a moderate ability to distinguish between the "change" and "no change" classes. Still, often the two classes are not well-separated, leading to a moderate-to-high rate of false positive and false negative detections across various thresholds. These errors were caused by the ML model prediction imprecision, which likely stemmed from the model training being influenced by noisy data. For example, Figure 4 shows two HISTARFM Landsat images in August (left) and September (right) 2020 for the California August Complex fire, which ignited on August 17, 2020. While the August image appears suitable for model training, the September image exhibits systematic horizontal striping artifacts across the entire image. With many other images with similar artifacts, the ConvLSTM model may overfit to the striping artifacts as false spatiotemporal signals, leading to reduced accuracy and generalizability. Moreover, the uncertainty bands failed to flag these artifacts as low-quality data, so even the weighted-loss model was unable to suppress their influence during model training.

6 Conclusion, Future Work, and Acknowledgment

In this paper, we presented a supervised ML method for detecting changes in forest canopy height using spatiotemporal HISTARFM Landsat and annual GEDI data at 30-meter resolution. While the method demonstrates moderate ability to distinguish between the "change" and "no change" classes, improvements are required to reduce false positives and false negatives before it can be deployed for real-world use cases. This is an ongoing effort, and we plan to address the identified problems from both data and modeling perspectives. To improve model robustness, we will incorporate Sentinel-1 dual-polarization SAR data (i.e., HH and HV) from the European Space Agency (ESA), which may provide complementary information in regions where Landsat optical data are affected by noise or cloud contamination. To mitigate the systematic underestimation of tall trees, we will investigate post-hoc correction techniques and explore adapting Focal Loss [8] for regression tasks to better balance the influence of tall versus short trees during training. The acknowledgment statement here has been anonymized for double-blind review.

References

[1] Elias Ayrey and Daniel J. Hayes. The use of three-dimensional convolutional neural networks to interpret lidar for forest inventory. *Remote sensing*, 10(4), 04 2018.

[2] Ralph Dubayah, James Bryan Blair, Scott Goetz, Lola Fatoyinbo, Matthew Hansen, Sean Healey, Michelle Hofton, George Hurtt, James Kellner, Scott Luthcke, John Armston, Hao Tang, Laura Duncanson, Steven Hancock, Patrick Jantz, Suzanne Marselis, Paul L. Patterson, Wenlu Qi, and Carlos Silva. The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of Remote Sensing*, 1, 06 2020.

[3] Ralph O. Dubayah and Jason B. Drake. Lidar remote sensing for forestry. *Journal of Forestry*, 98(6):44–46, 06 2000.

[4] Krzysztof M Gorski, Eric Hivon, Anthony J Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2), 2005.

[5] Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina, Peter V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. Humid tropical forest disturbance alerts using landsat data. *Environmental research letters*, 11(3), 03 2016.

[6] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8), 1997.

[7] Michael A. Lefsky, Warren B. Cohen, Geoffrey G. Parker, and David J. Harding. Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience*, 52(1):19–30, 01 2002.

[8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision*, 2017.

[9] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566, 02 2019.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.

[11] Fabian D Schneider, Felix Morsdorf, Bernhard Schmid, Owen L Petchey, Andreas Hueni, David S Schimel, and Michael E Schaepman. Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nat Commun*, 8(1), 11 2017.

[12] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai kin Wong, and Wang chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

[13] Marc Simard, Naiara Pinto, Joshua B. Fisher, and Alessandro Baccini. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116(G4), 11 2011.

[14] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, and Feng Xu. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing*, 5(4), 12 2017.

[15] Álvaro Moreno-Martínez, Emma Izquierdo-Verdiguier, Marco P. Maneta, Gustau Camps-Valls, Nathaniel Robinson, Jordi Muñoz-Marí, Fernando Sedano, Nicholas Clinton, and Steven W. Running. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. *Remote Sensing of Environment*, 247, 09 2020.