A 0.6-meter resolution canopy height and structure model for the contiguous United States

Scott L. Morford[1]*, Brady W. Allred[1], Shea P. Coons[1], Anthony A. Marcozzi[2] , Sarah E. McCord[3], Joseph T. Smith[1], David E. Naugle[4]

[1]Numerical Terradynamic Simulation Group, University of Montana, Missoula, MT, USA
[2]New Mexico Consortium, Los Alamos, NM
[3]Jornada Experimental Range, USDA Agricultural Research Service, Las Cruces, NM, USA
[4]W.A. Franke College of Forestry and Conservation, University of Montana, Missoula MT

Above-ground vertical structure is a critical variable for ecosystem monitoring, carbon accounting, and land management. However, the high cost and limited coverage of airborne lidar hinder its widespread application. To address this, we developed NAIP-CHM, a 0.6-meter resolution canopy height and structure model (CHM) covering the contiguous United States, derived from National Agriculture Imagery Program (NAIP) aerial imagery. Unlike forestry-specific models that exclude human-made features, NAIP-CHM characterizes the full vertical structure of the landscape including vegetation, buildings, and infrastructure. We utilized a U-Net convolutional neural network with attention mechanisms and environmental conditioning, training the model on 18 million lidar-derived training pairs with stratified sampling to ensure robustness in open-canopy ecosystems. Assessing performance against 2.3 million independent samples, the model achieved a root mean square error of 1.57 meters and coefficient of determination (r^2) of 0.87. We release the dataset, source code, and cloud-based tools to enable broad application without requiring specialized computational resources.

Background & Summary

Accurate characterization of ecosystem structure is essential for addressing foundational questions in ecology, carbon science, and land management[1-5] . While airborne and terrestrial lidar systems provide the most precise models of above-ground structure, their high acquisition cost constrains large-scale or repeated application. In contrast, public aerial and satellite imagery programs such as the National Agricultural Imagery Program (NAIP), Sentinel-2, and Landsat offer systematic, recurring coverage. When coupled with deep learning, these public resources provide a cost-effective pathway to produce spatially continuous canopy height models (CHMs).

Recent deep learning applications have produced several notable canopy height products. Global-scale mapping at 10-meter resolution has been demonstrated using Sentinel-2 and Global Ecosystem Dynamics Investigation (GEDI) lidar data[6] and meter-scale global models have been developed from commercial satellite imagery[7]. Regional efforts have further confirmed the utility of high-resolution aerial imagery for fine-scale mapping of canopy structure[8–10]. However, the accessibility of these products is often restricted by proprietary and costly inputs, closed-source code, or specialized expertise required to implement complex models. Furthermore, model development has historically prioritized forested ecosystems, often degrading performance in open-canopy environments like rangelands. Critically, many products mask human-made structures to support carbon accounting, limiting their utility for applications like wildfire modeling and wildlife habitat assessments where the physical presence of both vegetation and infrastructure matters. This represents a significant information gap, as open-canopy ecosystems and the wildland-urban interface are globally significant and central to land management[11,12].

To address these limitations, we introduce a 0.6-meter resolution canopy height and structure dataset for the contiguous United States (CONUS). Functionally acting as a normalized Digital Surface Model

**Corresponding author*: scott.morford@umontana.edu

(nDSM) by capturing all elevated features, the dataset is designated NAIP-CHM to highlight its focus on ecosystem structure. We generated the dataset using NAIP imagery spanning 2012 to 2023, with 96% of the imagery acquired in 2022–2023. We employed a U-Net[13] convolutional neural network with a Convolutional Block Attention Module[14] (CBAM) to refine feature maps and Feature-wise Linear Modulation[15] (FiLM) to integrate auxiliary environmental data. The model was trained using a public dataset of over 22 million NAIP and lidar derived CHM pairs[16].

We prioritize open access by providing the database through multiple channels suited to different user needs. For free bulk access independent of commercial cloud platforms, Cloud Optimized GeoTIFFS (COGs) are available for direct download from our servers. For users preferring cloud-based workflows, the data are also accessible via Google Earth Engine and a Google Cloud Storage bucket. To facilitate reproducibility and downstream application, we release the complete modeling pipeline, including the trained model, inference scripts, and a Google Colab notebook for applying the model to past and future NAIP imagery. Additionally, an interactive Google Earth Engine application is provided for no-code data exploration. These combined resources ensure the NAIP-CHM dataset is accessible to users with diverse technical expertise and computational capacity.

Methods

Model Training Data: NAIP Imagery and Lidar-Derived Canopy Heights

The model was trained using a public dataset of co-registered NAIP imagery and lidar-derived canopy height models (CHMs), comprising over 22 million CHM-NAIP pairs across CONUS[16]. The dataset is stratified by EPA Level III ecoregions[17] and National Land Cover Database[18] (NLCD) classes. To improve model performance in rangeland ecosystems, grassland and shrubland classes were oversampled by a factor of four and pasture by a factor of two.

The NAIP imagery is 4-band (R, G, B, NIR) orthoimagery at 0.6 or 0.3-meter resolution. Reference CHMs were generated from USGS 3D Elevation Program (3DEP) lidar point clouds using a standard pit-free algorithm[19] at 1.0-meter resolution. For model training, NAIP and reference CHMs were resampled to 0.6-meter. Co-located 432×432 pixel tiles were then extracted from the aligned NAIP and resampled CHM data, creating training pairs that each cover an area of approximately 256×256 meters.

Auxiliary Training Data

Auxiliary data were integrated to provide environmental context for model conditioning (Table 1). Climate variables were derived from a Principal Component Analysis (PCA) of 20 biologically-relevant features from 30-year PRISM[20] climatological normals (1991-2020); the first six principal components, capturing over 90% of the variance, were used. Soil properties were similarly summarized using the first six principal components from a PCA of over 20 variables from the SOLUS database[21], including texture, bulk density, and soil organic carbon. Topography was represented by a digital elevation model from the Shuttle Radar Topography Mission[22]. The day-of-year of each NAIP acquisition was encoded as sine and cosine components to account for phenology. Categorical context was provided by a simplified nine-class version of the 2021 NLCD[18] and EPA Level III ecoregion codes[17]. In total, 15 continuous and 2 categorical variables were used for model conditioning.

Model Architecture

The model is a U-Net[13] convolutional neural network enhanced with attention mechanisms and multi-stage FiLM[15] conditioning (Fig. 1). The model has 22,205,019 parameters and processes two data streams: the 4-band NAIP imagery and the auxiliary environmental features.

The U-Net backbone uses an encoder-decoder structure with channel dimensions of [64, 128, 256, 512, 1024] and skip connections. Each convolutional block consists of two 3×3 convolutions with stride 1 and padding 1, followed by Group Normalization (32 groups) and ReLU activation. These blocks are augmented with CBAM14 that applies sequential channel and spatial attention. Downsampling between encoder

Table 1: Datasets and Source Repositories Used For the NAIP-CHM Model Development

Data	Description	Resolution	Source
Primary Datasets			
NAIP Imagery	4-band (RGB-IR) orthorectified aerial imagery	0.6 m	a
LiDAR CHM	Canopy Height Models derived from 3DEP LiDAR	Resampled to 0.6 m	b
Conditioning Datasets			
PRISM	PCA of 30-year climatological normals (1991-2020)	~ 1km	c
SOLUS	PCA of Gridded soil properties database	100 m	d
SRTM	Shuttle Radar Topography Mission DEM	resampled to 240 m	e
NLCD	2021 National Land Cover Database classification	resampled to 240 m	f
EPA Ecoregions	Level 3 ecoregion classification	N/A (vector)	g

a) https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ;
b) http://rangeland.ntsg.umt.edu/data/rap/chm-naip/;
c) https://developers.google.com/earth-engine/datasets/catalog/OREGONSTATE_PRISM_Norm91m;
d) https://storage.googleapis.com/solus100pub/index.html;
e) https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003;
f) https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_NLCD;
g) https://developers.google.com/earth-engine/datasets/catalog/EPA_Ecoregions_2013_L3

stages is performed via 2×2 max pooling with stride 2, while upsampling in the decoder uses bilinear interpolation with a scale factor of 2. The encoder employs progressive spatial dropout (rates: 0.1 to 0.2) for regularization. The final output is generated by a 1×1 convolution that maps the decoded features to a single-channel canopy height map.

Auxiliary features are processed in a separate pathway. Categorical features (land cover, ecoregion) are converted to learned embeddings (8 and 16 dimensions, respectively). These are concatenated with the 15 continuous features (climate, soil, elevation, tem-

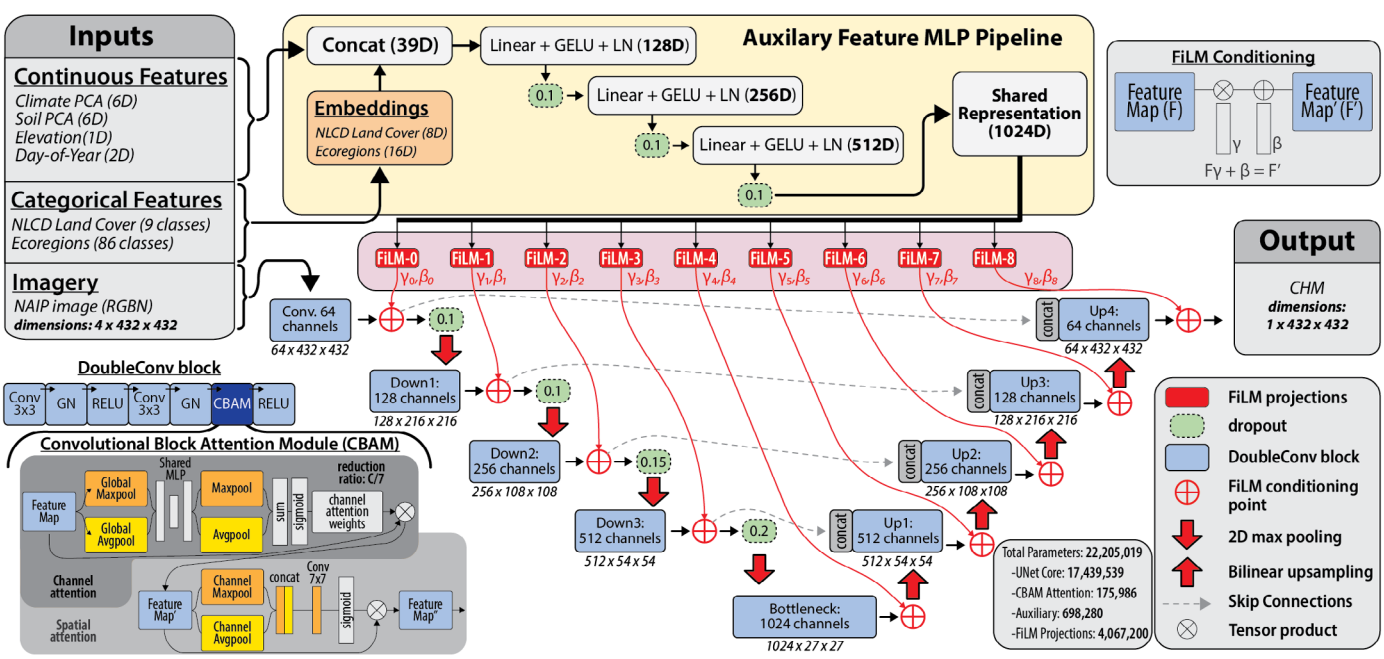

Figure 1: Schematic overview of the architecture used for Canopy Height Model (CHM) prediction. The network fuses high-resolution NAIP imagery with auxiliary environmental variables (e.g., climate, soil, land cover, ecoregion). The auxiliary features are processed via an MLP pipeline to generate a global representation, which dynamically modulates a U-Net backbone via Feature-wise Linear Modulation (FiLM) layers. The U-Net encoder-decoder path is further enhanced with Convolutional Block Attention Modules (CBAM) to refine spatial and channel-wise feature extraction.

poral). This 39-dimensional vector is processed by a three-layer multilayer perceptron employing GELU activation and Layer Normalization to produce a 1024-dimensional shared representation.

The FiLM conditioning mechanism modulates U-Net feature maps at nine locations (after the initial convolution and after each encoder and decoder stage). At each location, the shared auxiliary representation is projected to generate channel-specific scale (γ) and shift (β) parameters, which apply affine transformations to the feature maps. This multi-stage conditioning allows the model to integrate environmental information across multiple spatial scales.

Model Training

Data were partitioned for training, validation, and test sets following the splits provided with the source dataset. The final dataset consisted of 18,240,258 training samples (Fig. 2), 2,278,587 validation samples, and 2,273,381 test samples. Test samples were held out entirely during model development and used only to compute final model performance metrics.

Training followed a three-phase transfer learning strategy:

1. A U-Net with CBAM was trained from scratch on the 4-band NAIP imagery.
2. The U-Net and CBAM weights were frozen, and only the auxiliary multilayer perceptron and FiLM projection layers were trained.
3. All model weights were unfrozen for end-to-end fine-tuning.

A Huber loss function (delta=1.0) and the AdamW optimizer (learning rate: 6.0e-5, weight decay: 1.0e-4) were used. The learning rate was managed by a ReduceLROnPlateau scheduler. The model was implemented in PyTorch (v2.8.0) and trained on paired NVIDIA H100 GPUs with mixed-precision optimization over approximately four weeks. Geospatial processing used Rasterio (v1.3.0) and PyProj (v3.6.0).

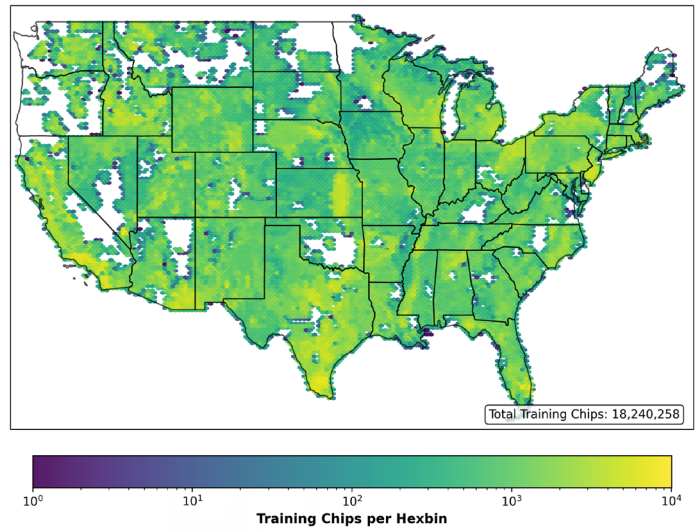

Figure 2: Geographic distribution of the paired NAIP and lidar derived canopy height model (CHM) training dataset. The map displays the spatial density of the 18,240,258 training chips used to develop the model. Hexbin colors represent the count of training samples per area on a logarithmic scale, covering the majority of the contiguous United States.

Model Evaluation

Test Dataset: Performance was evaluated on an independent test set with no spatial overlap with train ing tiles (n = 2,273,381). Test samples were filtered to remove tiles with sensor errors or data quality artifacts. This was achieved by analyzing reference height distributions to identify and exclude samples with insufficient valid pixel coverage or unrealistic height values and height distributions.

Evaluation Metrics: We computed mean absolute error (MAE), relative mean absolute error (RMAE), root mean square error (RMSE), bias, and the coefficient of determination (r^2) using standard methods. In addition to these standard metrics, we quantified the similarity between predicted and observed height distributions using the Jensen–Shannon Divergence (JSD), a symmetric measure of distributional agreement:

$$JSD(P \parallel Q) = \tfrac{1}{2} D_{KL}\big(P \parallel M\big) + \tfrac{1}{2} D_{KL}\big(Q \parallel M\big)$$

where

$$M = \frac{1}{2}\left(P + Q\right)$$

where P and Q are the normalized histograms of predicted and observed heights, respectively, M is their average distribution, and DKL denotes the Kullback-Leibler divergence:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right)$$

Height distributions were discretized into 60 bins spanning 0-60m (1m bin width), with a small constant ($\varepsilon = 10^{-10}$) added to prevent numerical instability. JSD values are bounded between 0 and 1, with lower values indicating greater agreement between predicted and observed distributions. This metric provides insight into whether the model captures the full variability in canopy height beyond point- and tile-wise accuracy.

Error Attribution: We modeled tile-level prediction error (log[MAE]) to evaluate the relative influence of observed canopy height, land cover class (NLCD groups), and Level III ecoregion on spatial variation in model performance. We first fit univariate models to identify the most informative observed canopy height metric (tile mean vs. 95th percentile). The selected height metric was then included as a fixed effect in a set of linear mixed-effects models with random intercepts for (1) NLCD class, (2) ecoregion, or (3) both NLCD class and ecoregion. From the best-supported model, we partitioned variance among fixed, random, and residual components to quantify the contribution of each factor. This analysis revealed how variation in model error reflects both geographic context (ecoregion and land cover) and scene structure, evaluating the model's ability to generalize across diverse environmental gradients in CONUS.

CONUS-wide Dataset Generation

Using the most recent NAIP imagery available, we produced a canopy height dataset across CONUS at 0.6-meter resolution (Fig. 3). We used NAIP Digital Ortho Quarter Quads (DOQQs) as the primary unit of inference. DOQQs were retrieved using Google Earth Engine[23] and split into 432x432 pixel tiles with 20% overlap. In all, we processed 213,567 DOQQs, covering approximately 7.7 million km^2. The resulting CHMs were saved as Cloud-Optimized GeoTIFFs (COGs) with DEFLATE compression and internal overviews. The complete dataset with accompanying metadata totals approximately 24.82 TB.

Data Records

The NAIP-CHM dataset, model, and associated code are distributed across several platforms to ensure accessibility and usability for different applications.

1. CONUS Canopy Height Dataset: The NAIP-CHM 0.6-meter dataset is publicly available through multiple access points. For bulk download, COGs and associated index files are available via HTTP from the Rangeland Analysis Platform server (*http://rangeland.ntsg.umt.edu/data/naip-chm/*). Access through Google Earth Engine (*ImageCollection ID: projects/naip-chm/assets/conus-structure-model*) facilitates interactive analysis, particularly for non-specialists. For programmatic cloud-based workflows, COGs are also available in a public Google Cloud Storage bucket (*gs://naip-chm-assets/*) with "requester pays" enabled. In addition to the data, we provide a simple, interactive web application for data visualization and exploration, found at *https://naip-chm.projects.earthengine.app/view/naip-chm-a-conus-structure-model*.

2. Model and Code Archive: A version-of-record archive of the model code, model weights, validation data, and auxiliary data used to generate the results presented in this manuscript is deposited in the Zenodo repository[24].

3. Development Repository and Code: The full codebase for model training, inference, and data processing is available in GitHub (https://github.com/smorf-ntsg/naip-chm). This repository includes examples demonstrating how to apply the model to future or past NAIP imagery.

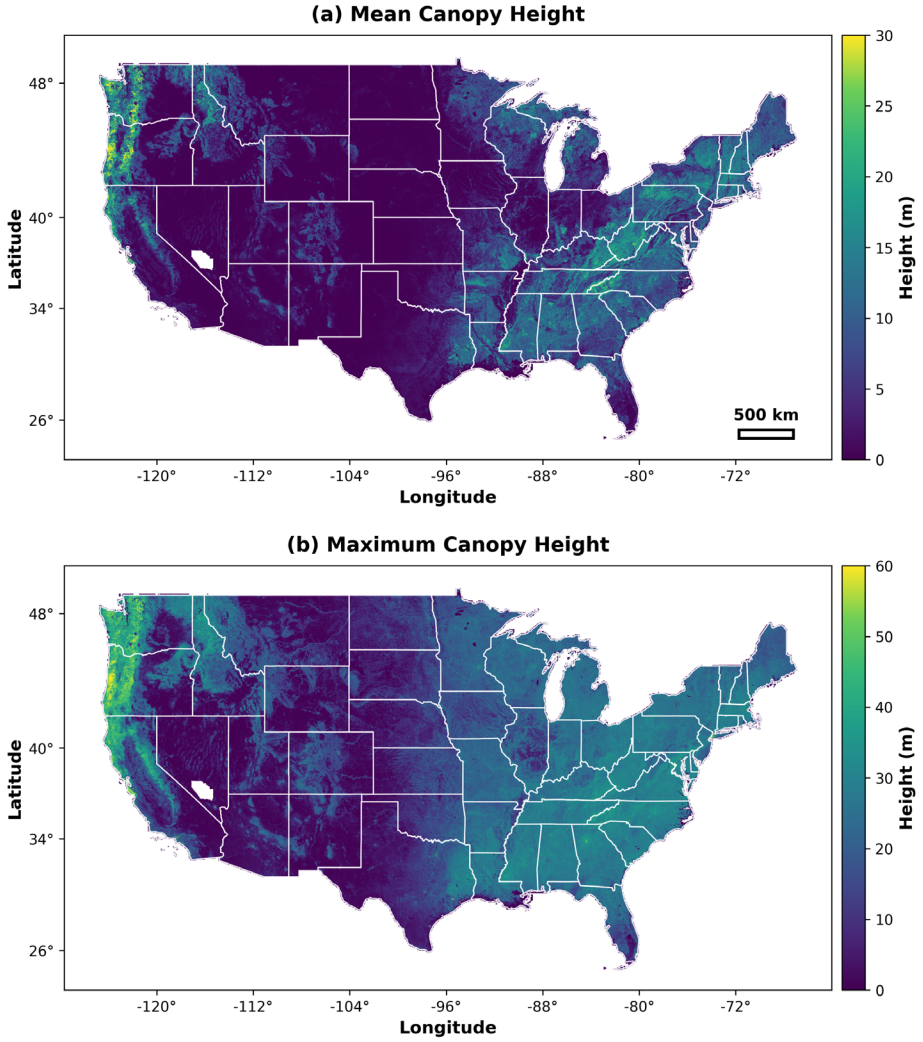

Figure 3: Modeled canopy height. Maps display the (a) mean and (b) maximum canopy height derived from the NAIP-CHM. The inset in (a) displays the distribution of acquisition years for the source NAIP imagery (y-axis is log-scale). To characterize patch-level structure, the maximum height in panel (b) represents the maximum of 38 m mean-aggregated patches, rather than native 0.6 m maxima. The void in coverage in southern Nevada is the Nevada Test and Training Range, where NAIP imagery is restricted.

Technical Validation

Overall, pixel-level mean absolute error (MAE) was 0.893 m, root mean square error (RMSE) was 1.574 m, and bias was -0.022 m (Table 2). The r^2 was 0.871, and the Jensen-Shannon Divergence (JSD) was 0.0107, indicating close correspondence between predicted and observed height distributions.

At the tile level (432x432 pixels, approximately 256 × 256-meter spatial resolution; n = 2,276,165), predicted mean and 95th percentile heights showed strong agreement with reference data (r^2 = 0.98 and 0.97, respectively) (Fig. 4a, b). The model exhibits a known compression effect, underestimating the tallest canopies by approximately 10%, a pattern consistent with similar deep learning methods[8–10]. Performance may be reduced in forests with canopies exceeding 70 m, such as those in the Pacific Northwest, as these areas were generally absent from the training data.

Absolute error increased with canopy height, whereas relative error decreased (Fig. 4c–e). The error partitioning analysis attributed 59.3% of the variance in prediction error to the 95th percentile canopy height. In contrast, the random effects of land cover and ecoregion combined accounted for only 11.8% of the variance (conditional r^2 = 0.71), indicating that model error is driven more by local canopy structure—specifically, underestimation of very tall canopies—than by broad

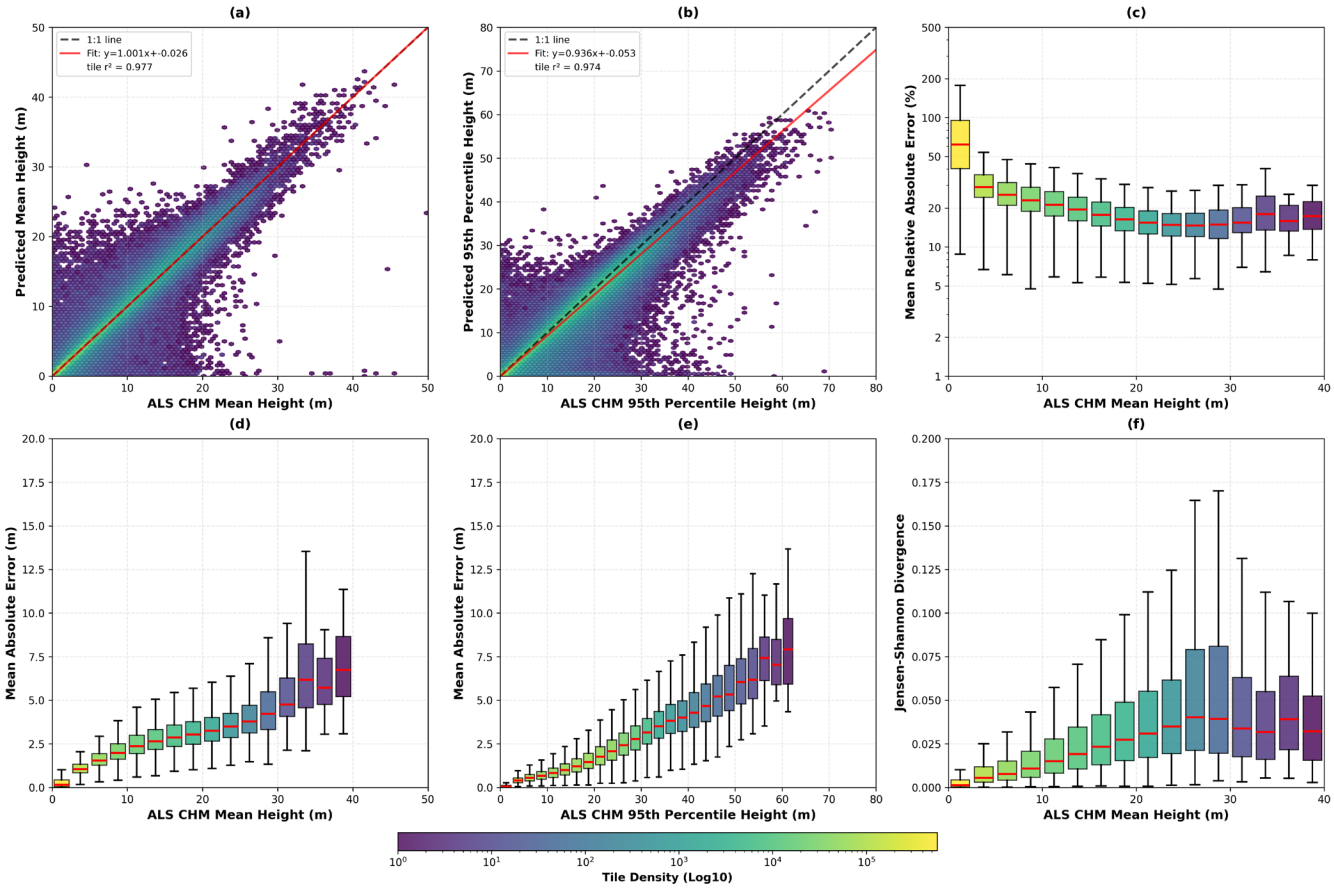

Figure 4: Tile-level validation of the NAIP derived Canopy Height Model (CHM) against the lidar derived CHM reference data. Panels (a) and (b) display density scatter plots for mean height and 95th percentile height, respectively, including the linear fit (red solid line) and 1:1 identity (black dashed line). Panels (c–f) present box plots of error metrics binned by reference height: (c) Mean Relative Absolute Error in log-scale, (d) Mean Absolute Error (MAE) versus mean height, (e) MAE versus 95th percentile height, and (f) Jensen-Shannon Divergence. Red lines within the box plots indicate the median value.

geographic or land-cover context. Analysis of the Jensen-Shannon Divergence shows that mean JSD

Table 2: Model performance metrics on the test dataset

MAE (m)	0.893
RMSE (m)	1.574
Bias (m)	-0.022
r^2	0.871
Jensen–Shannon divergence	0.0107
Pixels	421.58 x 10^9
Tile Count	2,276,165

Note: Note: All performance metrics except Jensen-Shannon divergence were computed on a pixel-by-pixel basis, not aggregated at the tile level. The original CHM reference data were at 1.0 m resolution but were resampled to 0.6 m for modeling.

increases only modestly with canopy height, confirming that the model accurately reproduces the reference height distributions within tiles across a range of canopy structures (Fig. 4f).

The geographic distribution of model performance reflects this error analysis (Fig. 5). Higher absolute errors (MAE) are concentrated in regions with tall canopies, such as the Pacific Northwest, northern Rocky Mountains, and Appalachians (Fig. 5a). Regional bias remains near zero, though a subtle pattern of positive bias in eastern forests and negative bias in western forests is apparent (Fig. 5b). r^2 values are high (>0.8) across most ecoregions, and JSD remains low (<0.05), confirming consistent performance (Fig. 5c, d).

For out of sample validation, NAIP-CHM was compared against CHMs from the National Ecologi-

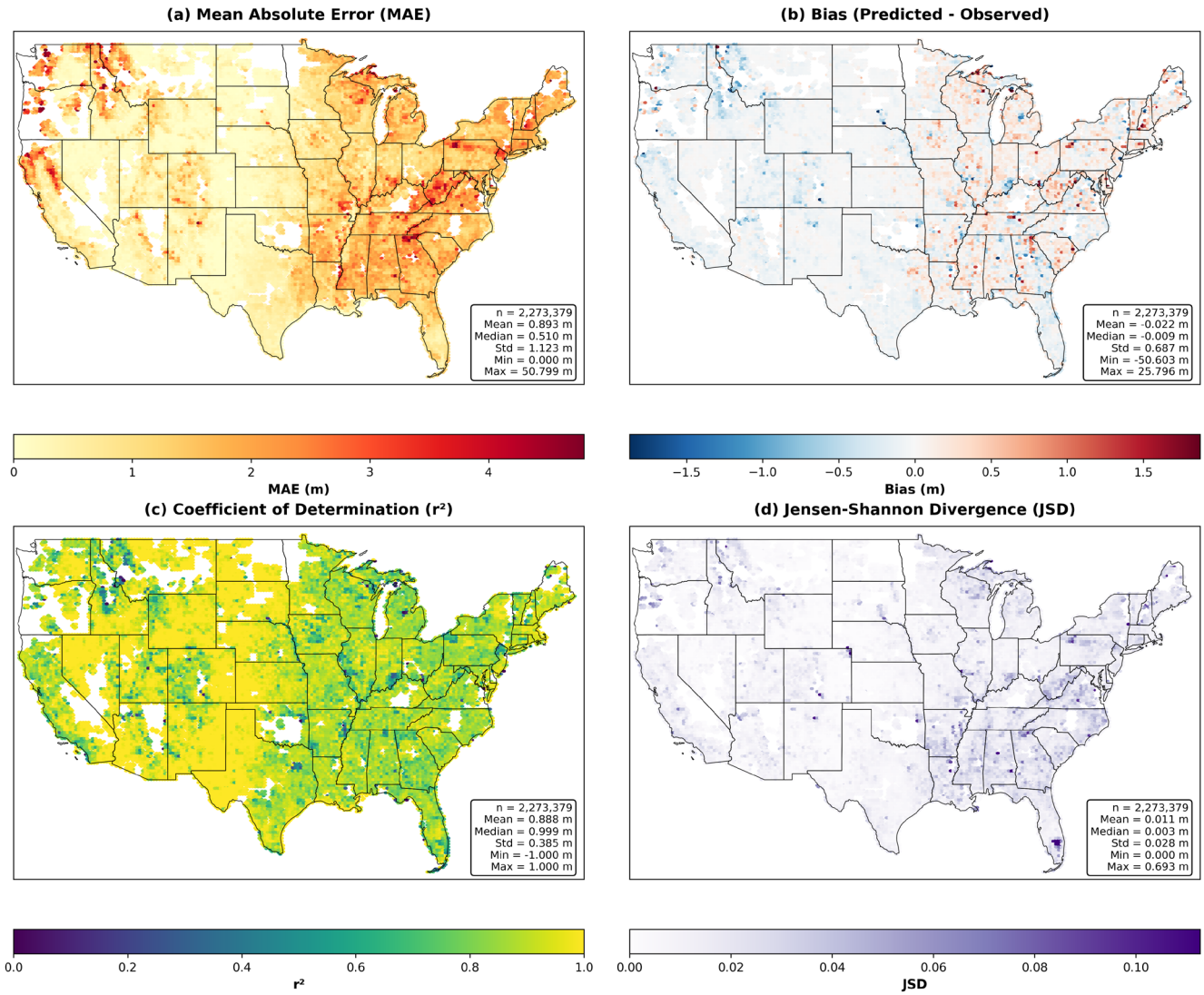

Figure 5: Geographic distribution of validation metrics across the contiguous United States. Panels display tile-level (a) Mean Absolute Error (MAE), (b) Bias, (c) Coefficient of Determination (r^2), and (d) Jensen-Shannon Divergence (JSD). Inset boxes provide global summary statistics for each metric.

cal Observatory Network (NEON) Airborne Observation Platform at seven sites (Fig. 6). Across six of the seven sites, MAE ranged from 0.58 to 3.24-meter and r^2 from 0.63 to 0.92. The seventh site (MOAB) was excluded from this summary due to unrealistic height values in the NEON reference CHM. It is important to note that lidar-derived CHMs are themselves models subject to processing artifacts and error, and are not absolute ground truth. The out of sample comparison with NEON sites confirms the model's performance and ability to produce comparable data.

Temporal Applications

NAIP imagery is generally collected on a 2-3 year cycle. Although the NAIP-CHM dataset produced herein relied on the most recent NAIP imagery to produce a single snapshot of canopy height and structure, the model can be applied to both past and future NAIP 4-band imagery. This enables high resolution monitoring, capturing structural changes after disruptions or change events (see Figure 7), highlighting the model's advantage for both broad-scale inventory and site-specific evaluation.

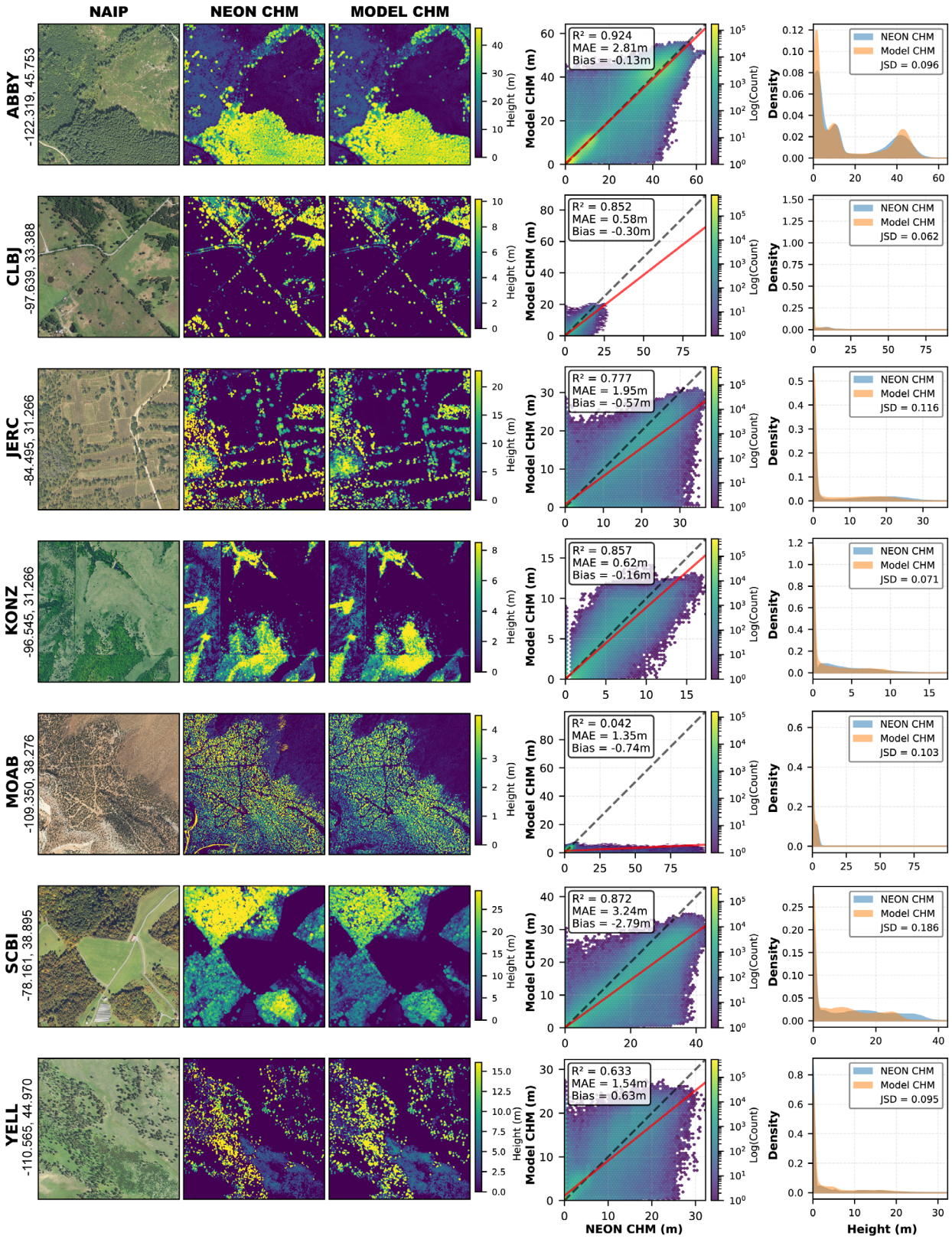

Figure 6: Out of sample validation using National Ecological Observatory Network (NEON) canopy height models (CHM) at seven sites. Scatter plots and histograms display performance metrics (r^2, MAE, Bias, and Jensen-Shannon Divergence). The poor performance at MOAB is due to artifacts and unrealistic values in the NEON CHM. The dashed black line represents the 1:1 line, while the red line indicates the regression fit.

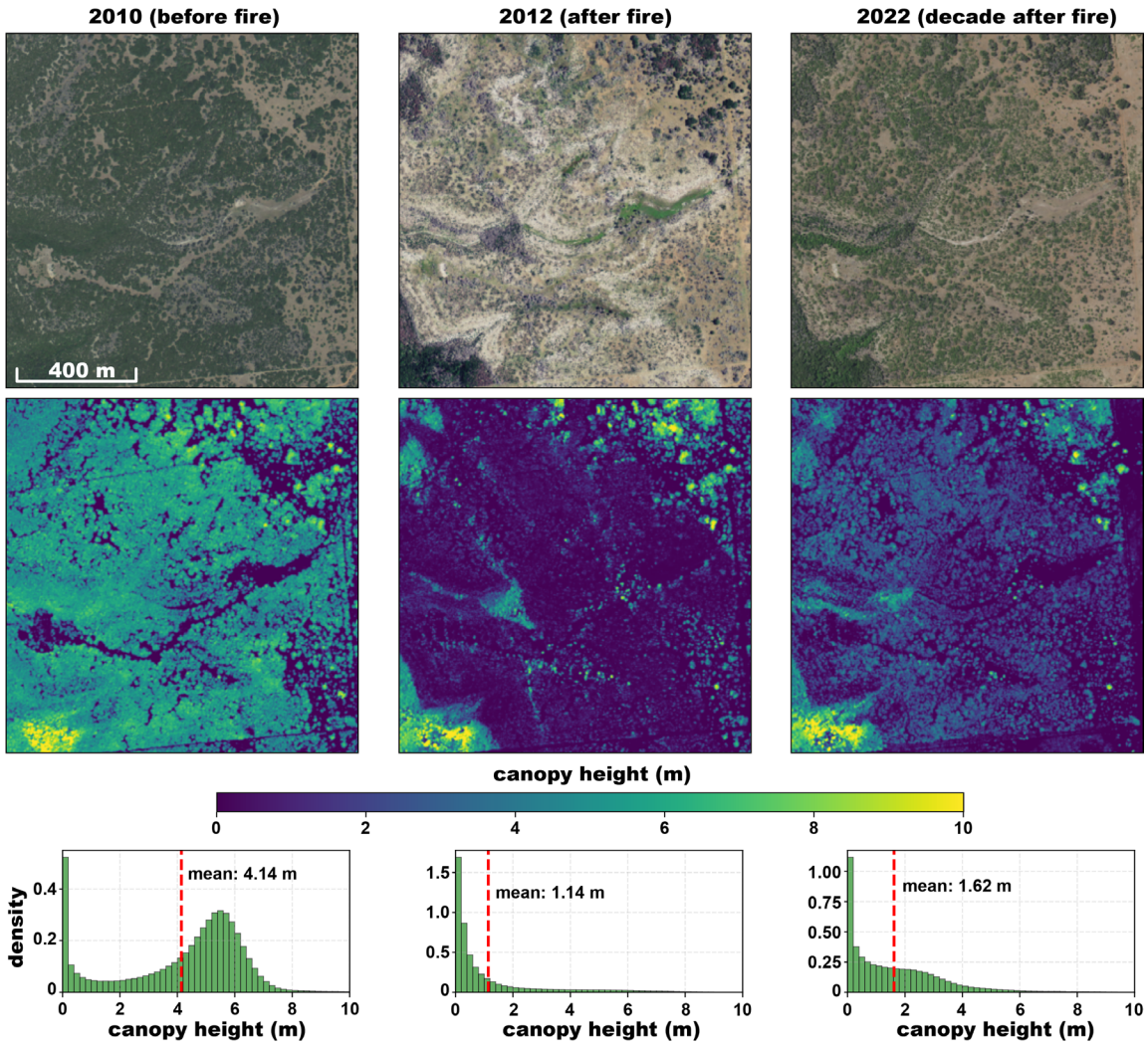

Figure 7: Chronosequence showing vegetation recovery following 2011 wildfire in central Texas (latitude 32.58, longitude -98.61). (a) Study area in 2010 (before fire), (b) 2012 (after 2011 fire), and (c) 2022 (~ decade after fire). Middle row: CHMs showing the loss of canopy structure immediately following the fire event and subsequent regeneration.The density distributions show the shift from a closed woodland height structure (2010) to predominantly low vegetation (2012) and partial recovery (2022).

Data Availability

The 0.6-meter resolution CONUS canopy height dataset is publicly available for bulk download via HTTP from the Rangeland Analysis Platform server (*http://rangeland.ntsg.umt.edu/data/naip-chm/*), as a Google Earth Engine asset (ImageCollection ID: projects/naip-chm/assets/conus-structure-model), and as Cloud-Optimized GeoTIFFs in Google Cloud Storage requestor-pays bucket (*gs://naip-chm-assets*).

Code Availability

The source code, trained model weights, validation data, and auxiliary datasets required to reproduce the results are permanently archived in a Zenodo repository[24]. The active development repository, which contains scripts for model training, inference, and validation, is hosted on GitHub (*https://github.com/smorf-ntsg/naip-chm*) under an MIT License. To facilitate application of the model without local computational resources, we provide a Google Colab notebook

(https://colab.research.google.com/github/smorf-ntsg/naip-chm/blob/main/notebooks/gee_inference_colab.ipynb) that allows users to run inference on past or future NAIP DOQQs.

Acknowledgements

We thank Kris Mueller for assistance in figure preparation. This research used compute resources provided by (1) the SCINet project and/or the AI Center of Excellence of the USDA Agricultural Research Service (ARS project numbers 0201-88888-003-000D and 0201-88888-002-000D); (2) Google Earth Engine; (3) the Numerical Terradynamic Simulation Group at the University of Montana; and (4) the University of Montana's Hellgate Research Cluster. Funding for this project was provided by USDA-NRCS (NR230325XXXXC009 to S.L.M.); Pheasants Forever (WLFW C008-2025-02 to D.E.N. and S.L.M.); and the DoD SERDP (RC20-1025, "Closing Gaps") and ESTCP (RC23-7626, "FastFuels Tool Suite") projects to A.A.M. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author Contributions

S.L.M. led the conceptualization, methodology, software development, data curation, formal analysis, and visualization (web application), and wrote the original draft. B.W.A. contributed to conceptualization, software (model inference, code review), and validation. A.A.M. contributed to validation and software (code review). J.T.S. contributed to the formal analysis. D.E.N. provided supervision. All authors reviewed the model outputs, provided feedback during model development, and contributed to the review and editing of the manuscript.

References

1. Dubayah, R. et al. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002 (2020).

2. Clawges, R., Vierling, K., Vierling, L. & Rowell, E. The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. Remote Sens. Environ. 112, 2064–2073 (2008).

3. Skowronski, N. S., Clark, K. L., Duveneck, M. & Hom, J. Three-dimensional canopy fuel loading predicted using upward and downward sensing LiDAR systems. Remote Sens. Environ. 115, 703–714 (2011).

4. Wulder, M. A. et al. Lidar sampling for large-area forest characterization: A review. Remote Sens. Environ. 121, 196–209 (2012).

5. Smith, W. K. et al. Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. Remote Sens. Environ. 233, 111401 (2019).

6. Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. A high-resolution canopy height model of the Earth. Nat. Ecol. Evol. 7, 1778–1789 (2023).

7. Tolan, J. et al. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. Remote Sens. Environ. 300, 113888 (2024).

8. Campbell, M. J., Eastburn, J. F., Brewer, S. C. & Dennison, P. E. Deep learning and machine learning enable broad-scale woodland height, cover, and biomass mapping. ISPRS J. Photogramm. Remote Sens. 226, 187–203 (2025).

9. Chang, T. et al. VibrantVS: A High-Resolution Vision Transformer for Forest Canopy Height Estimation. Remote Sens. 17, 1017 (2025).

10. Wagner, F. H. et al. Sub-meter tree height mapping of California using aerial images and LiDAR-informed U-Net model. Remote Sens. Environ. 305, 114099 (2024).

11. Reiner, F. et al. More than one quarter of Africa's tree cover is found outside areas previously classified as forest. Nat. Commun. 14, 2258 (2023).

12. Morford, S. L. et al. Herbaceous production lost to tree encroachment in United States rangelands. J. Appl. Ecol. 59, (2022).

13. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Preprint at https://doi.org/10.48550/arXiv.1505.04597 (2015).

14. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. CBAM: Convolutional Block Attention Module. in 3–19 (2018).

15. Perez, E., Strub, F., Vries, H. de, Dumoulin, V. & Courville, A. FiLM: Visual Reasoning with a General Conditioning Layer. Proc. AAAI Conf. Artif. Intell. 32, (2018).

16. Allred, B. W., McCord, S. E. & Morford, S. L. Canopy height model and NAIP imagery pairs across CONUS. Sci. Data 12, 322 (2025).

17. Omernik, J. M. & Griffith, G. E. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. Environ. Manage. 54, 1249–1266 (2014).

18. Jon Dewitz. National Land Cover Database (NLCD) 2021 Products. U.S. Geological Survey https://doi.org/10.5066/P9JZ7AO3 (2023).

19. Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T. & Hussin, Y. A. Generating Pit-free Canopy Height Models from Airborne Lidar. Photogramm. Eng. Remote Sens. 80, 863–872 (2014).

20. Daly, C. et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064 (2008).

21. Nauman, T. W. et al. Soil landscapes of the United States (SOLUS): Developing predictive soil property maps of the conterminous United States using hybrid training sets. Soil Sci. Soc. Am. J. 88, 2046–2065 (2024).

22. Farr, T. G. et al. The Shuttle Radar Topography Mission. Rev. Geophys. 45, (2007).

23. Gorelick, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27 (2017).

24. Morford, S.L. NAIP-CHM: A 0.6-meter resolution canopy height and structure model for the contiguous United States. (2025). Zenodo. https://doi.org/10.5281/zenodo.17664995