USDA Forest Service
U.S. DEPARTMENT OF AGRICULTURE

Geospatial Office | GO-10268-RPT2 | September 2025

National Land Cover Database Tree Canopy Cover Methods

Version: 2023.5

Mapping Areas: Conterminous United States; Southeastern Alaska; Hawaii; and Puerto Rico and U.S. Virgin Islands

Cover image: Tree canopy cover for the conterminous United States.

Housman, I.W.; Heyer, J.P.; Ruefenacht, B.; Schleeweis, K.; Megown, K.; Bogle, S; Reischmann, J.; Ryerson, D. 2025. National Land Cover Database Tree Canopy Cover Methods v2023.5. GO-10268-RPT2. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Field Services and Innovation Center- Geospatial Office. 31 p.

Executive Summary

The National Land Cover Database (NLCD) Tree Canopy Cover (TCC) product is a remote sensing-based map output produced by the United States Department of Agriculture, Forest Service (Forest Service). The 2023 NLCD TCC release includes annual maps from 1985 to 2023 across the conterminous United States (CONUS), southeast Alaska (SEAK), Puerto Rico and the US Virgin Islands (PRUSVI), and Hawaii. Study areas outside CONUS (OCONUS) are expected to be released in summer 2025. TCC Science products are also available annually from 1985 to 2023. Science and NLCD TCC products can be used in many applications including monitoring forest health, post-disturbance recovery, urban canopy distribution, and carbon accounting.

This document details the methods used to create the annual Science and NLCD TCC map products for TCC version 2023.5. A forthcoming manuscript will provide a more comprehensive summary of TCC 2023.5 background, methods, and results. These methods will be revisited with each new release to ensure they reflect the best available science. Current methods utilize Google Earth Engine (GEE) to run the Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) temporal segmentation algorithm on annual composites of Landsat and Sentinel-2 imagery. The outputs from LandTrendr are used as predictor variables in random forest models that are calibrated outside GEE using Forest Service Forest Inventory and Analysis (FIA) photo-interpreted TCC data to make wall-to wall TCC predictions on a pixel-wise basis.

The main goal of the 2021.4 version of the methods was to enable more frequent time steps, a longer temporal extent, and more coherent temporal signals in the TCC products. The 2023.5 version extends these goals further by mapping 1985 to 2023. These goals were achieved by using GEE storage and compute capacity for image processing and incorporating time series-based approaches already utilized by Forest Service programs such as the Landscape Change Monitoring System (LCMS), while also ensuring the NLCD TCC product retains the level of quality expected of a product that is part of the NLCD suite.

NLCD TCC products are freely available for download from the Multi-Resolution Land Characteristics Consortium (MRLC).

Annual Science TCC products are available for download from the Forest Service GeoData Clearinghouse.

Authors

Ian Housman (technical lead) is a contractor employed by RedCastle Resources as a senior remote sensing specialist in the Field Services and Innovation Center–Geospatial Office (FSIC–GO).

Josh Heyer (production lead) is a contractor employed by RedCastle Resources as a geospatial specialist in the FSIC–GO.

Bonnie Ruefenacht (production lead) is a contractor employed by RedCastle Resources as a geospatial programmer in the FSIC–GO.

Karen Schleeweis is the science team lead at the Rocky Mountain Research Station, Forest Inventory and Analysis.

Kevin Megown is the program leader for the Resource Mapping, Inventory, and Monitoring (RMIM) program in the FSIC–GO.

Seth Bogle is a contractor employed by RedCastle Resources as a geospatial project manager.

Jaclyn Reischmann is a geospatial data services specialist for RMIM in the FSIC–GO.

Daniel Ryerson was a geospatial specialist for RMIM in the FSIC–GO.

Acknowledgements

The authors would like to thank Mark Finco, Wendy Goetz, and Vicky Johnson for their contributions to the methodological development of the TCC product.

Contents

Tree Canopy Cover Conterminous United States, Southeast Alaska, Puerto Rico and the U.S. Virgin Islands, Hawaii version 2023.5 Release Notes

Any changes to the methods from Tree Canopy Cover (TCC) version 2021.4 outlined below in this document will be reflected in this list.

- Computing platforms
  - No changes
- Model Calibration Data
  - No changes
- Model predictor data
  - Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) was rerun with annual composite data from 1984-2024.
- Modeling
  - No changes were made.
- Map Assemblage
  - Different Tau percentiles were used for low density and high density urban TCC.
  - Different Tau percentiles were used for the Landscape Change Monitoring System (LCMS) Land Cover Tall Shrub class in southeast Alaska.
  - Different Tau percentiles were used for the LCMS Land Cover Shrub class in Hawaii.
  - Different Tau percentiles were used for the LCMS Land Cover Non-Processing Area class in OCONUS.
  - We did not apply a clump and eliminate minimum-mapping unit to non-urban TCC.
- TCC Products
  - We will release v2023.5 TCC for southeast Alaska, Puerto Rico–U.S. Virgin Islands, and Hawaii in late summer 2025/early Fall 2025.

Background

The United States Department of Agriculture (USDA), Forest Service (Forest Service) produces the National Land Cover Database (NLCD) Tree Canopy Cover (TCC) product as a member of the Multi-Resolution Land Characteristics (MRLC) consortium (Wickham et al., 2014). The 2001 version of the NLCD TCC product was produced by the United States Geological Survey (USGS; Homer et al., 2004). All subsequent releases have been produced by the Forest Service. This includes the 2011, 2016, v2021.4 releases, and this most recent v2023.5 release. With each new NLCD TCC data release, previous datasets are superseded with the newest version for each map year. This follows protocols used for the NLCD Land Cover and Urban Impervious Cover datasets.

Annual nationwide mapping programs are becoming more common (Brown et al., 2020; USDA Forest Service 2025). Annual NLCD TCC products and annual TCC Science products (raw direct model outputs and per pixel model standard deviation, referred to as standard error) extending from 1985-2023 are being released. To allow for easy tracking of the product being used, all products and documentation share a version naming convention: YYYY.v. “YYYY” denotes the most recent year mapped, and “v” denotes the version of the methods.

Production methods are collaboratively developed by the NLCD TCC Science Team under the Forest Service Forest Inventory and Analysis (FIA) program and the NLCD TCC Production Team at the Field Services and Innovation Center–Geospatial Office (FSIC–GO). Methods are reviewed and revised as necessary with each new release. A primary objective of the 2021.4 release was to integrate the Forest Service Landscape Change Monitoring System (LCMS) methods and data (USDA Forest Service 2025). Consequently, NLCD TCC and LCMS now share much of the same data and image processing workflows. Because NLCD TCC v2023.5 shares several methods with the v2021.4 release and LCMS, this document will pull from the v2021.4 methods documentation (Housman et al., 2023) as well as the LCMS methods documentation (Housman et al., 2022) where appropriate.

NLCD TCC product suite covers the conterminous United States (CONUS), southeast Alaska (SEAK), Puerto Rico and US Virgin Islands (PRUSVI), and Hawaii. CONUS will be released in spring 2025. Study areas outside CONUS (OCONUS) will be released in late summer/early fall 2025. The NLCD TCC product suite includes 30 m spatial resolution maps of tree canopy cover density and annual science products from 1985-2023. This document describes methods for CONUS. Methods for OCONUS are still being developed. This document will be updated when OCONUS methods are finalized.

Study Areas

Figure 1 shows the coverage of the v2023.5 product suite. All study areas except SEAK match those of USGS produced NLCD datasets. Specifically, CONUS consists of the 48 contiguous states. Hawaii is limited to the eight major islands: Ni'hau, Kaua'i, O'ahu, Moloka'i, Lāna'i, Kaho'olawe, Maui, and Hawai'i. The Puerto Rico portion of PRUSVI consists of Puerto Rico, Culebra, Vieques, and Mona. The U.S. Virgin Islands portion of PRUSVI consists of St. Croix, St. Thomas, and St. John. SEAK includes a 219,000 km^2 region of southeast and south-central coastal Alaska (Figure 1).

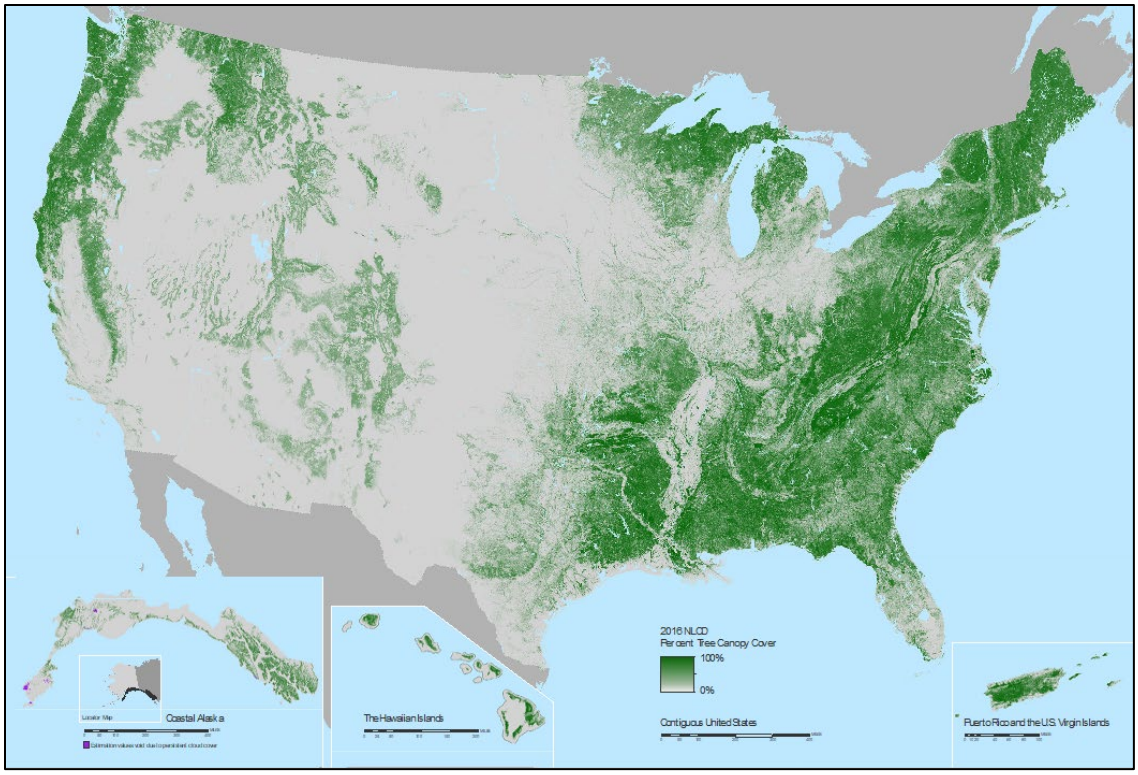

Figure 1. Depiction of the tree canopy cover study areas for the conterminous United States, southeast Alaska, Hawaii, and Puerto Rico and the U.S. Virgin Islands.

Methods

Computing Platforms

NLCD TCC utilizes Google Earth Engine (GEE; Gorelick et al., 2017) for all remote sensing raster data acquisition and processing through an enterprise agreement between the Forest Service and Google. GEE is a parallel computing environment that provides access to many publicly available earth observation datasets, common data processing methods, and computing infrastructure.

To maintain confidentiality of the reference data locations collected over the FIA sample grid (Bechtold & Patterson 2005), we used R (R Core Team 2020) and the Python package Scikit-Learn (Pedregosa et al., 2011) on local compute resources at FSIC–GO to extract predictor data over reference locations, build random forest TCC models, and compute model and map error.

Reference Data

Reference Data Collection

Reference data were collected using high-resolution imagery and the Canopy Cover Tool, developed by the Forest Service as an ArcMap™ extension (Figure 2; Goeking et al., 2012). To compute a plot's estimated tree canopy cover percentage, a grid of 109 dots (with a spacing of 26.25 feet) is placed within a 144-foot radius circle centered on the plot coordinates. The 109 dots are rotated 15 degrees to avoid following linear anthropomorphic features in cardinal directions. Photo interpreters indicate whether the dot falls upon tree canopy or not and the percent tree

canopy cover is calculated for the plot. Interpreters also assign confidence levels in their photo interpretations for each plot. The plot coordinates used for reference data collection were located on the FIA grid (Reams et al., 2005).

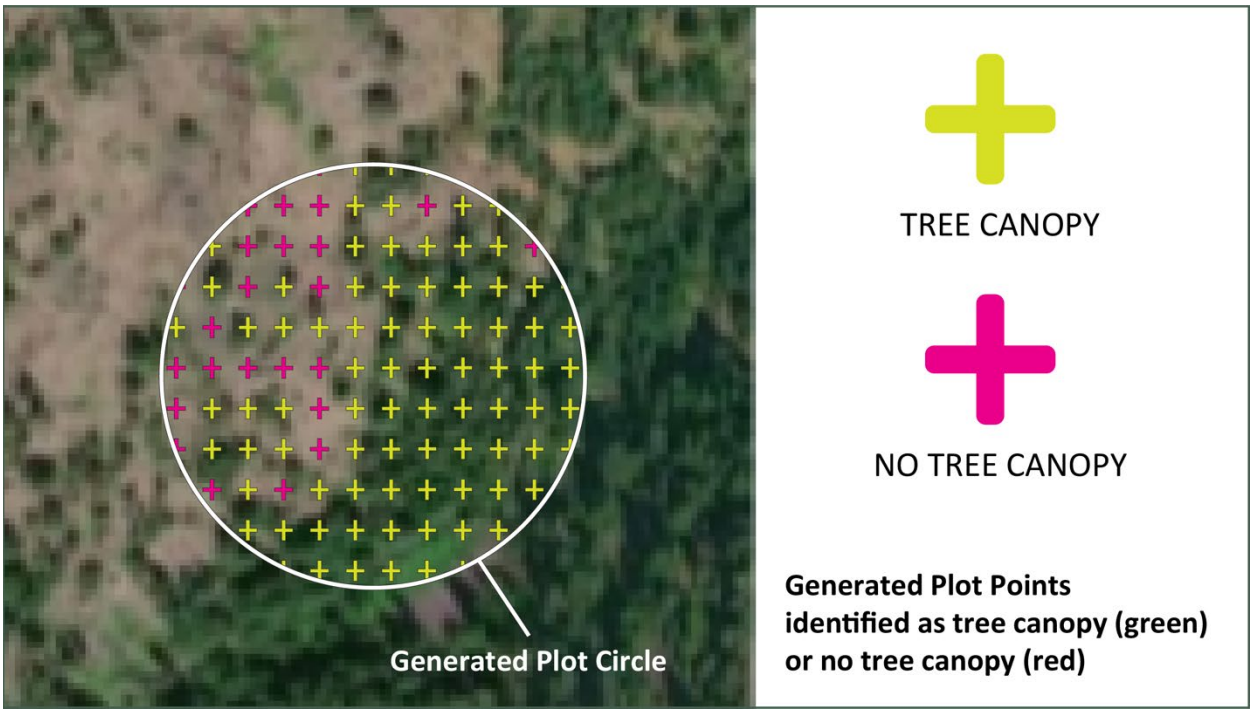

Figure 2. Depiction of the Canopy Cover tool used to develop tree canopy cover reference data.

For this project, tree canopy is defined as:

- Tree – live tally trees, saplings and seedlings – where tally trees are all species listed here: https://www.fia.fs.usda.gov/library/field-guides-methods-proc/index.php. This list is updated and archived according to the FIA field guide.
- Canopy – The vertically-projected polygon described by the outline of the foliage, ignoring any normal spaces occurring between the leaves of plants, and ignoring overlap among multiple layers of a species (Daubenmire 1959).

Percent tree canopy cover is then defined as:

- The percent of systematically arranged points (109 total) within a 144-foot radius circle centered on an FIA plot that fall over tree canopy as defined above.

Reference Data Filtering

To minimize error in our reference data, we omitted plots for the following reasons:

- Low interpreter confidence
- Spectrally unstable (possible change)
- LCMS Change (Loss or Gain)
- Plots located in the ocean
- No cloud or cloud shadow-free spectral information available

- Plots photo-interpreted as greater than zero TCC, but rechecks indicated TCC equal to zero. Because of the ambiguity of these plots, they were discarded.

Of the 63,010 FIA plots, 55,356 were included as reference data (Table 1). These plots were randomly divided into model calibration (70 percent) and map error assessment (30 percent) groups.

Table 1. Forest Inventory and Analysis (FIA) plots were removed from conterminous U.S. (CONUS) reference data due to low photo-interpretation (PI) quality scores, duplicate plots, locations over oceans or unavailable data, unstable spectral data, inaccurate zero canopy PI, and PI that spectrally changed according to Landscape Change Monitoring System (LCMS) change data between 2009-2012 (within 1 year of NAIP PI acquisition window). A total of 7,654 PI plots were removed from reference data, leaving 55,356 PI observations across CONUS available for training and validation.

FIA PI reason for removal	Number of PI plots removed	PI plot count
All FIA Plots		63,010
Removed previously	476	
Low photo-interpretation quality scores	1,972	
Duplicates	703	
Ocean or unavailable spectral data	20	
Spectrally unstable	47	
QA of zero canopy PI called canopy	260	
LCMS Change (Loss or Gain) 2009-2012	4,176	
Final number of PI plots used		55,356

Model Predictor Data

Our model predictor data consisted of Landsat and Sentinel-2 imagery fitted by the Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) temporal segmentation algorithm, topographic information from the United States Geological Survey (USGS) 3-Dimensional Elevation Product (3DEP), and the USDA National Agricultural Statistics Service Cropland Data Layer (CDL). Descriptions of each of these datasets are provided below.

Remote Sensing Spectral Data

LandTrendr Fitted Composites

We used Landtrendr fitted composites for the predictor spectral dataset. The sections below describe the creation of Landtrendr fitted composites.

Data Preparation

We used USGS Collection 2 Tier 1 Level 1 Landsat 4, 5, 7, 8, and 9 and Sentinel-2a and 2b level 1C top of atmosphere reflectance data, as available in GEE fall 2024. We did not use surface reflectance data because the Sentinel-2 surface reflectance data available within GEE are terrain-corrected. This makes it difficult to utilize in unison with Landsat surface reflectance data that are not terrain-corrected.

For cloud masking Landsat data, we applied the CFmask cloud and cloud shadow masking algorithm (Foga et al., 2017), which is an implementation of Fmask 2.0 (Zhu & Woodcock 2012), as

well as the cloudScore algorithm (Chastain et al., 2019). For cloud masking Sentinel-2 data, we utilized the s2Cloudless algorithm outputs available through GEE (Zupanc 2020). We masked cloud shadows in both Landsat and Sentinel-2 using the Temporal Dark Outlier Mask (TDOM) method (Chastain et al., 2019). All remote sensing data preparation procedures are available in the FSIC–GO GEE data processing and visualization library:

- Python Module on PyPI – https://pypi.org/project/geeViz/
- Python Module on GitHub – https://github.com/gee-community/geeViz

Annual Compositing

We utilized annual composites of the masked Landsat and Sentinel-2 data as inputs for LandTrendr. Annual composite values are the geometric medoid of all values not masked as cloud or cloud shadow from a specified date range for each year. Due to differences in data availability and seasonality, we adapted the date window range across different modeling regions (Table 2).

Table 2. Dates used for annual compositing of Landsat and Sentinel-2 data

Study Area	Pre-Sentinel-2 Start Date	Pre-Sentinel-2 End Date	Post-Sentinel-2 Start Date	Post-Sentinel-2 End Date
Conterminous U.S.	June 1	September 30	July 1	September 1
Southeastern Alaska	June 15	September 15	June 15	September 15
Puerto Rico – U.S. Virgin Islands	June 1	May 31	June1	May 31
Hawaii	January 1	December 31	January 1	December 31

The geometric medoid is the value that minimizes the sum of the square difference between the median value of each band's values (Flood 2013). This ensures that the center-most value in a multi-dimensional feature space is chosen. The value from all bands is from the same observation date. The bands that we included in the feature space are green, red, near infrared (NIR), first shortwave infrared (SWIR1), and second shortwave infrared (SWIR2). We omitted the blue band because it is more prone to atmospheric scattering and can inappropriately influence the medoid algorithm. Any pixel that did not have a cloud or cloud shadow free value for a given year was left as null and excluded from the composite for that year. The 2020 composite images for CONUS, SEAK, Hawaii, and PRUSVI are shown in Figure 3, Figure 4, and Figure 5.

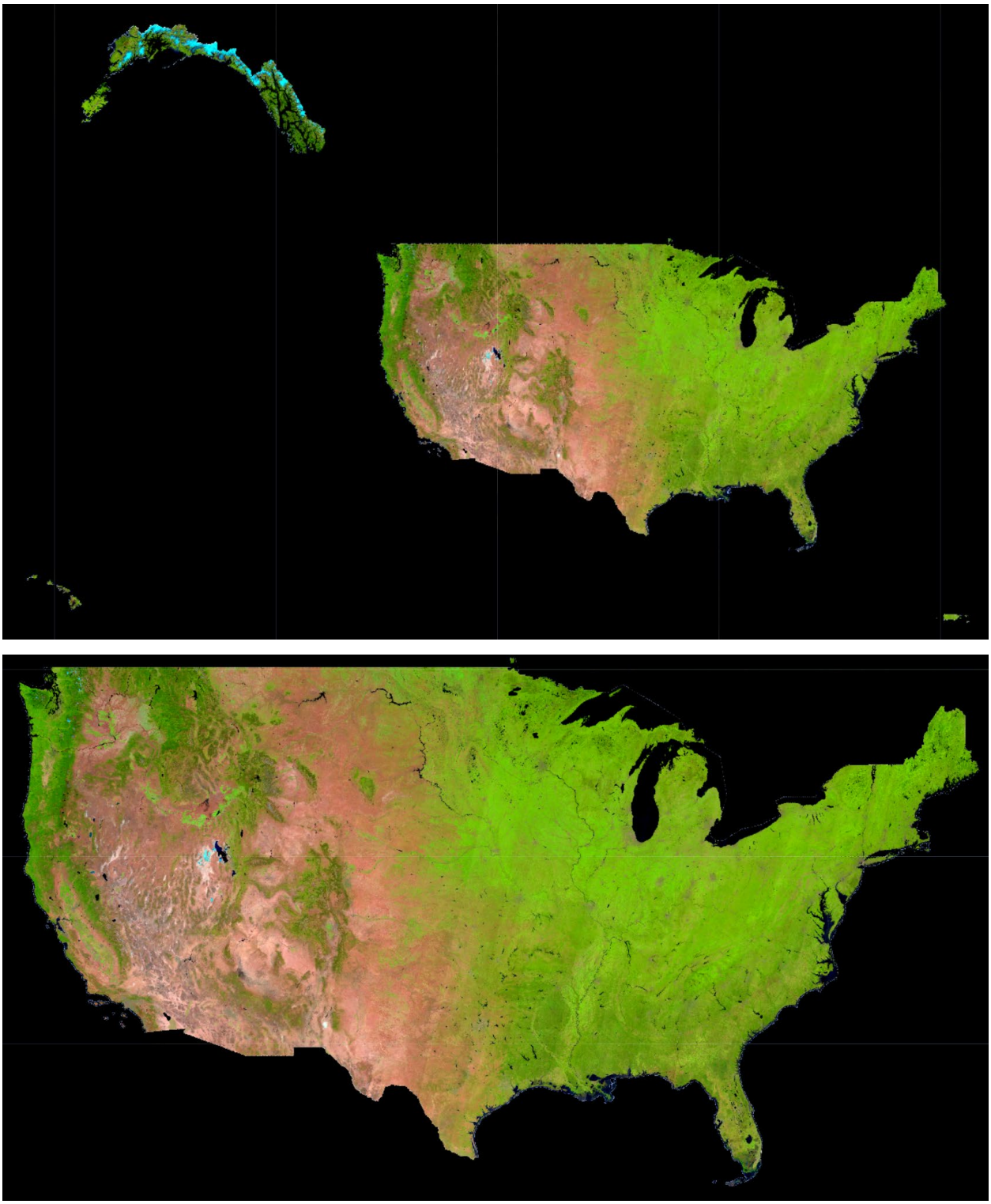

Figure 3. Example of the 2020 composites used in the National Land Cover Database Tree Canopy Cover product. The top image includes all study areas, and the bottom image shows the conterminous U.S. The red, green, and blue channels used in these composites are the second shortwave infrared (SWIR2), near infrared (NIR), and red bands, respectively.

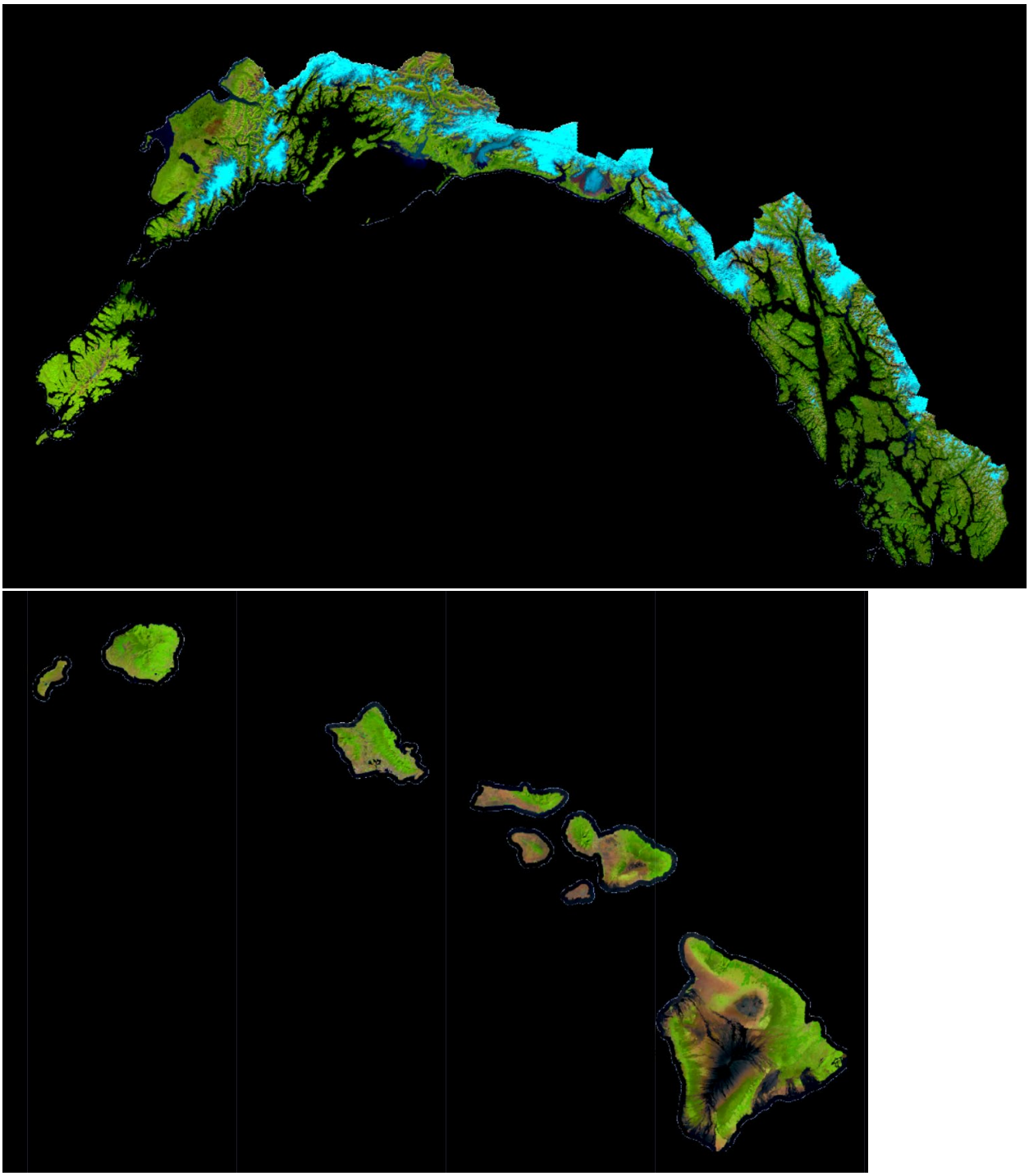

Figure 4. Example of the 2020 composites used in the National Land Cover Database Tree Canopy Cover product. The top image shows Southeastern Alaska, and the bottom image shows Hawaii. The red, green, and blue channels used in these composites are the second shortwave infrared (SWIR2), near infrared (NIR), and red bands, respectively.

Figure 5. Example of the 2020 composites used in the National Land Cover Database Tree Canopy Cover product . This image shows Puerto Rico and the U.S. Virgin Islands. The red, green, and blue channels used in these composites are the second shortwave infrared (SWIR2), near infrared (NIR), and red bands, respectively.

With the workflow outlined above, two separate sets of annual composite series from 1984 to 2024 were created. The first set includes Landsat 7 from 1999-2002 when the scanline corrector worked properly. In May 2003, the Landsat 7 scanline corrector failed, introducing stripes of missing data throughout all images. This first set does not include any Landsat 7 data with scanline correction striping. The second set utilizes Landsat 7 from 1999-2015 both with and without the scanline correction striping artifact. To minimize scanline correction striping artifacts, only SWIR1, SWIR2, normalized difference vegetation index (NDVI), , the normalized burn ratio (NBR), normalized difference moisture index (NDMI), and tasseled cap brightness are used. Note that the first set is different from those used for LCMS and it is used for CONUS, coastal AK, and PRUSVI study areas. Hawaii lacked sufficient Landsat 5 imagery to create composites from 2003-2013 without Landsat 7.

Temporal Segmentation

The goal of temporal segmentation is to identify periods of time that likely have similar vegetation cover and/or change processes. We used LandTrendr (Kennedy et al., 2010; Kennedy et al., 2018) to segment both sets of prepared composite image time series. LandTrendr requires a maximum of one observation per year (i.e., an annual composite made from Landsat and Sentinel-2 data).

LandTrendr Methods

LandTrendr performs iterative temporal segmentation of annual composite time series, yielding a set of segments. Each segment is characterized by a start year, end year, and associated fitted values at the segment's start and end vertices (Figure 6). Based upon this information, fitted values for each band or index for each year were used to generate the LandTrendr fitted composites used as predictor layers.

Figure 7 illustrates how one of the predictor variables, NBR, was segmented by LandTrendr for two pixels. The pixel depicted in the left figure shows a fire event, while the right figure shows insect-related tree mortality. The dark green line is the raw composite NBR value, while the lighter green line represents the fitted values we used in our models.

We implemented the Landtrendr methods as described above using the GEE version of LandTrendr (Kennedy et al., 2018). The parameters used were the same as described in Kennedy et al., (2018) (Table 3). The 2011 fitted LandTrendr values were used for model calibration. All years were used for model application.

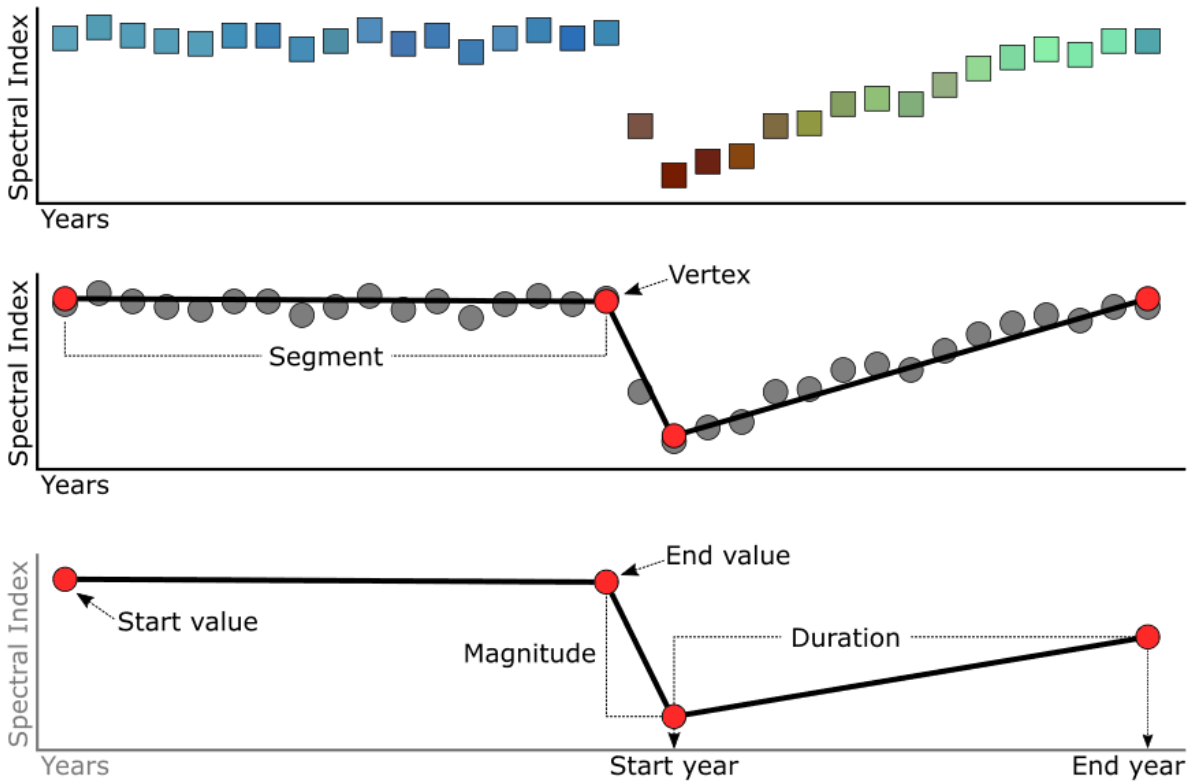

Figure 6. Illustration from https://emapr.github.io/LT-GEE/ depicting how Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) breaks a time series and the information that can be taken from the output.

Figure 7. An example of raw medoid composite and Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) fitted normalized burn ratio (NBR) values for a single pixel. The pixel depicted in the left chart shows a fire event, while the right chart shows insect-related tree mortality.

Table 3. Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) parameters and descriptions.

Parameter Name	Value	Description
maxSegments	6	Maximum number of segments to be fitted on the time series.
spikeThreshold	0.9	Threshold for damping the spikes (1.0 means no dampening).
vertexCountOvershoot	3	The initial model can overshoot the maxSegments + 1 vertices by this amount. Later, it will be pruned down to maxSegments + 1.
preventOneYearRecovery	true	Prevent segments that represent one-year recoveries.
recoveryThreshold	0.25	If a segment has a recovery rate faster than 1/recoveryThreshold (in years), then the segment is disallowed.
pvalThreshold	0.05	If the p-value of the fitted model exceeds this threshold, then the current model is discarded and another one is fitted using the Levenberg-Marquardt optimizer.
bestModelProportion	1.25	Takes the model with most vertices that has a p-value that is at most this proportion away from the model with lowest p-value.

Terrain Data

We used terrain metrics, derived from 10 m USGS 3DEP data (U.S. Geological Survey 2019) to provide elevation, slope, and aspect, information to the model. The specific variables used were:

- Elevation
- Aspect
- Sine of aspect
- Cosine of aspect
- Slope

We resampled this dataset from 10 m to 30 m spatial resolution using cubic convolution. All the other topographic variables were derived from the resampled elevation dataset.

Cropland Data

We used the NLCD land cover cultivated crops and pasture-hay classes (U.S. Geological Survey 2024) and the CDL cropland layers (Lin et al., 2022; USDA National Agricultural Statistics Service 2008) to provide information on agricultural areas. For each model year, binary agriculture data were produced by classifying non-tree CDL crops as agriculture and everything else as non-agriculture.

Summary

Table 4 shows a complete list of predictor variables considered for modeling.

Table 4. List of the National Land Cover Database Tree Canopy Cover product model predictor variables. This table depicts annual spectral predictors, which are different for each year of the analysis period, and static terrain predictors, which remain constant through time. Sets 1 and Sets 2 differ in the inclusion or exclusion of Landsat 7 bands. Set 1 excluded Landsat 7 bands after 2002, but did use all Landsat 4, Landsat 5, Landsat 8, Landsat 9, Sentinel 2a, and Sentinel 2b when those satellites were available. Set 2 included Landsat 7 after 2002, but excluded the visible bands, near-infrared band, and the normalized difference vegetation index for Landsat 7 until 2015 when Sentinel 2a replaced Landsat 7.

Predictor Variables	**Set 1:** **(No Landsat 7 after 2002)**	**Set 2:** **(Includes Landsat 7 until 2015)**
Spectral Bands		
blue	✓	
green	✓	
red	✓	
near infrared	✓	
shortwave infrared 1	✓	✓
shortwave infrared 2	✓	✓
Indices/Transformations		
normalized difference vegetation index	✓	
normalized burn ratio	✓	✓
normalized difference moisture index	✓	✓
normalized difference snow index	✓	✓
brightness	✓	✓
greenness	✓	
wetness	✓	
brightness/ greenness angle	✓	✓
Agriculture		
cropland data layer	✓	✓
Terrain (static)		
elevation	✓	✓
slope	✓	✓
cosine of aspect	✓	✓
sine of aspect	✓	✓

Modeling

NLCD TCC utilized the random forest modeling method (Breiman 2001), which randomly selects a subset of predictor variables and reference data creating many different regression trees. Each of the trees predicts a TCC value, the means of which are used on a pixel-wise basis as the final modeled TCC values.

We utilized the GEE instance of random forests called “smileRandomForest” for all raster-based regression. To meet confidentiality requirements of the FIA reference data locations, all predictor data extraction and model calibration were performed locally using the sklearn.ensemble.RandomForestClassifier method. Calibrated models were then uploaded to GEE for application to the raster predictor stack.

Due to the size and wide variety of CONUS ecotones, modeling was broken up into 54 480x480 km tiles. To negate possible seamlines between tiles and to improve model accuracies by increasing sample sizes, we expanded the single tile modeling areas to 5x5 grids of tiles (Figure 8). Model calibration plots that intersected corresponding 5x5 grids were combined with data from 2011 LandTrendr fitted composites, 2011 CDL, and terrain data. Random forest models were built using these model calibration datasets representing 5x5 grids of tiles and the models were applied to only the center tiles of the 5x5 grids. Figure 8 illustrates how the moving window approach covered CONUS.

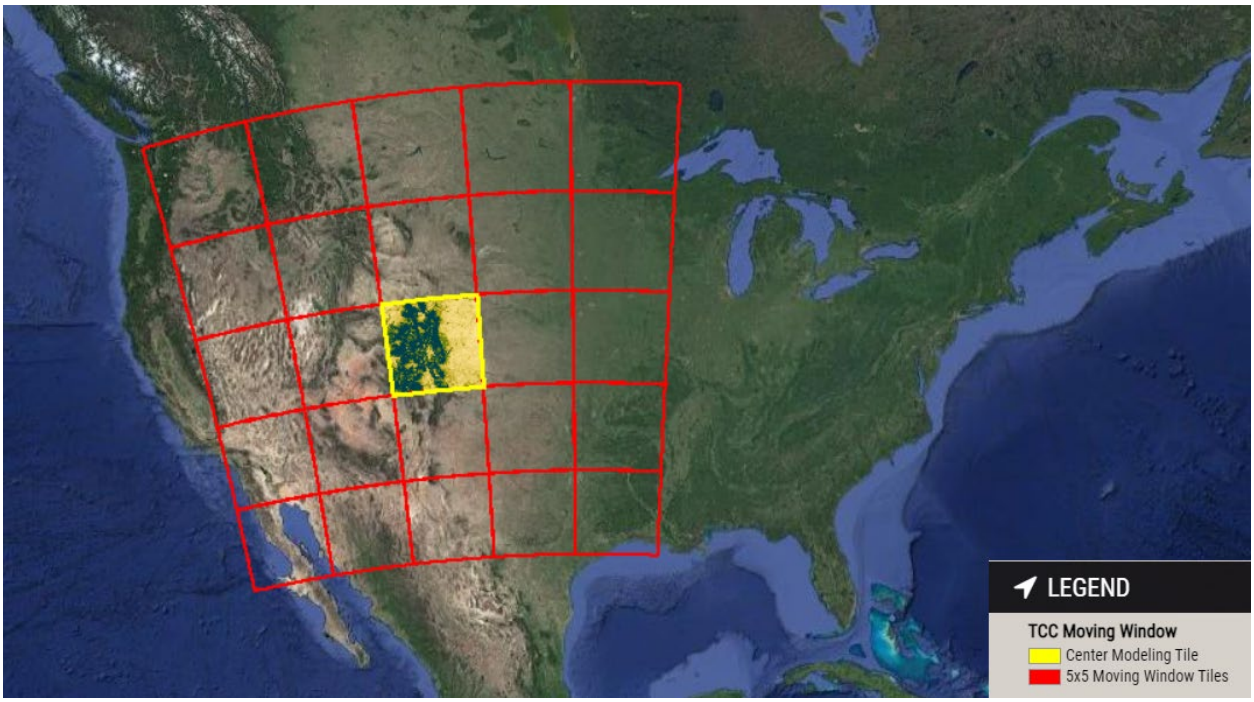

Figure 8. Illustration of 5x5 480x480 km moving window modeling method. This frame illustrates the adjacent tiles in red where model calibration data were pulled from, and the yellow tile in the center is where the model was applied. The raw Science 2011 Tree Canopy Cover (TCC) map is depicted as it is predicted for this tile. Note: this image uses a web Mercator projection which is not an equal area projection, so the equal area square tiles appear distorted. An animated illustration of this moving window method is available here: https://storage.googleapis.com/tcc-graphics/TCC_Tile_Animation.gif.

These models were built on relationships between 2011 reference data and 2011 predictor data. Temporal fitting and segmenting algorithms (i.e., LandTrendr) vastly improve the signal to noise ratio in time-series built with Landsat and Sentinel 2 data. It has generally been accepted that radiometric normalization enables the spatial and temporal transfer of models, built with a limited subset of training data, to predict over additional years or locations that were not used in model calibration (Hansen & Loveland 2012; Moisen et al., 2016; Potapov et al., 2022; Powell et al., 2010; Vogeler et al., 2018). Following this logic, we applied the 2011 model to predict TCC from 1985 to 2023. The last year of the image time series fitted by LandTrendr (2024) was not used to predict, as the last year can have more errors.

While each pixel has a mean predicted TCC value for each year, it also has a standard deviation of the predicted values from all regression trees; we refer to this as standard error. The utility of the standard error is limited since it does not account for the actual model error or the variability of that model error. To provide a depiction of the variability of expected model error, for each of the model calibration datasets for the 54 tiles, we simulated a model error statistic referred to as tau (τ) (Equation 1), following guidance from Coulston et al., (2016).

$$\text{Equation 1} \qquad \tau = \sqrt{\frac{(Observed - Predicted)^2}{var(Predicted)}}$$

The intent of τ is to provide a measure of model uncertainty. The implementation is designed to normalize pixel-wise standard error by the simulated variance of model error. Bootstrap resampling is used to parameterize many random forest models, where for each model the observed and predicted values are retained for the ~ 37 percent holdout (typical in bootstrap sampling with replacement). Those held out pairs of observations and predictions represent a type of error assessment that can be used to define the prediction and error of new observations. In theory, the distribution of these pairwise bootstrapped holdouts can be used to build intervals to help evaluate the uncertainty or confidence around future predictions in the TCC distribution space. These τ distributions are translated to tables and provided as part of the Science TCC products.

Science Products

Science products consist of the mean and standard deviation (also referred to as standard error throughout the documentation of this release) of the TCC values (Figure 9), along with the τ tables for each of the 54 models. These outputs are available annually from 1985-2023 through the Forest Service TCC distribution website (https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/index.php).

This output is useful for anyone who prefers to introduce their own post-processing workflow and/or needs a depiction of pixel-wise model error and model uncertainty.

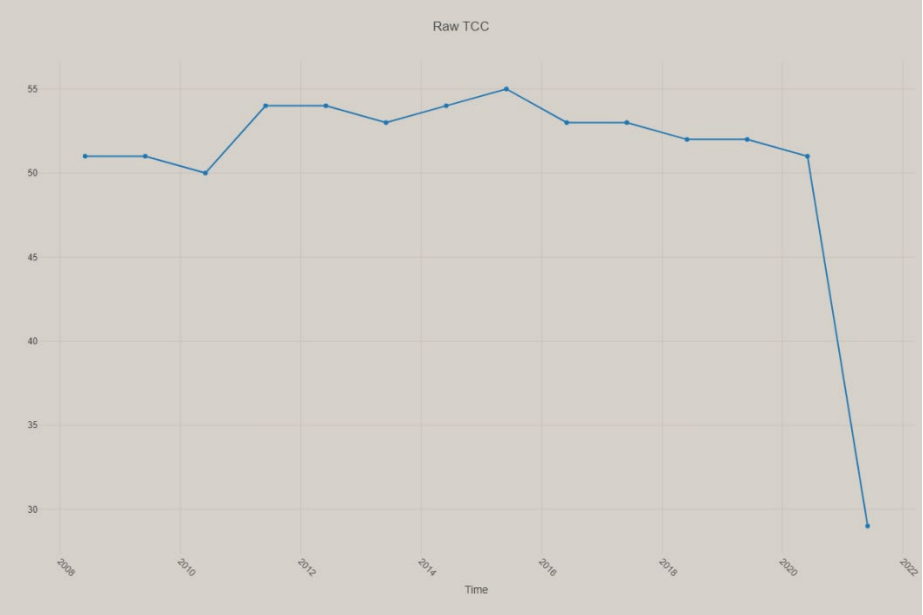

Figure 9. Example of a pixel of the raw Science Tree Canopy Cover (TCC) output. The values will bounce around a little each year. This pixel was initially a healthy forest, followed by a long period of gradual decline of TCC from 2015-2020. Then in late summer 2020 (2021 mapped year), this area experienced a wildfire.

NLCD TCC Post-processing Methods

We developed a post-processing workflow to create the final NLCD TCC products. The main workflow steps were:

- For non-urban areas:
  - mask non-treed features and set TCC value to zero percent
- For urban areas:
  - use the τ statistic to mask areas with low TCC and high model uncertainty
- For all mapping areas, remove small interannual TCC changes from the annual TCC time series to reduce noise and smooth the time series.

Separating Urban and Non-Urban Areas

To mitigate masking highly fragmented TCC common in urban settings, we separated urban and non-urban areas. We used 2018 TIGER U.S. Census Block data (U.S. Census Bureau, Geography Division 2024) and the LCMS land use developed class as our urban mask.

Non-urban TCC Masking

For non-urban areas, we created annual non-tree masks to mitigate errors related to false positive where predicted TCC was greater than zero but most likely was zero. For each year, we used the following masks:

- LCMS land cover tree class 3 year moving window (+/- 1 year from year of interest) mode to create tree mask (USDA Forest Service 2025)
- NLCD land cover cultivated crops and pasture-hay class to create non-tree crop mask (U.S. Geological Survey 2024)

- CDL tree crop classes to unmask tree crops as TCC (USDA National Agricultural Statistics Service 2008)
- For CONUS, the NLCD land cover water class was used to create water mask (U.S. Geological Survey 2024)
- For OCONUS, the annual LCMS Land Cover product water class was used to create a water mask (USDA Forest Service 2025)

We recoded TCC to zero for pixels that were flagged as non-treed in the non-tree masks (Figure 10; Figure 11).

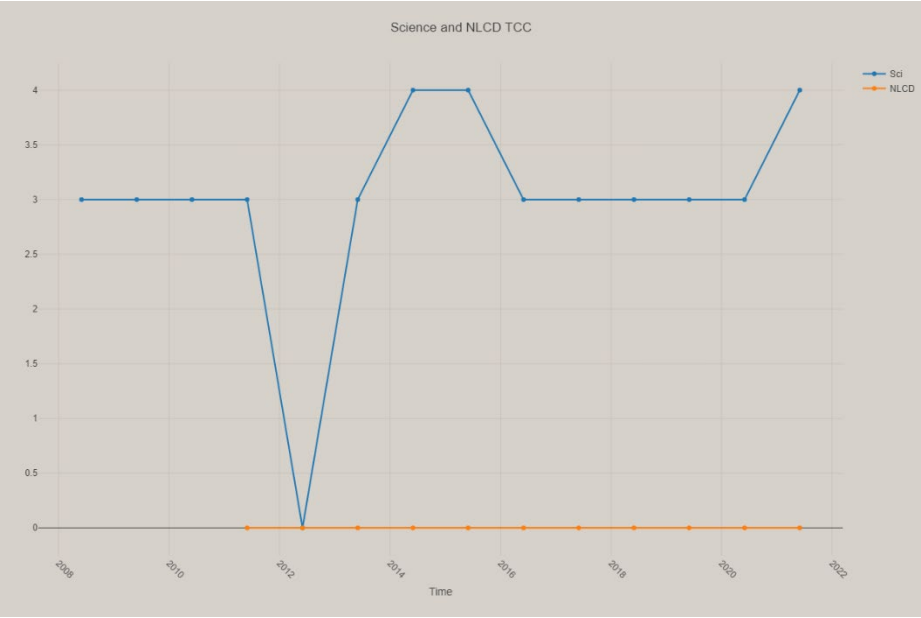

Figure 10. An example pixel where the non-urban non-tree mask filter was applied. As all years fell within the non-urban non-tree mask, the Tree Canopy Cover (TCC) values in the Science TCC product were set to 0 in the National Land Cover Database (NLCD) TCC product.

For SEAK and Hawaii, trees found in the LCMS Land Cover tall shrub class and shrub classes were omitted from the initial LCMS non-tree masks. The 87th τ percentile was used to separate treed from non-treed pixels in the LCMS Landcover tall shrub class in SEAK. The 91st τ percentile was used to separate treed from non-treed pixels in the LCMS Landcover shrub class in Hawaii. For tree masking in LCMS shrub classes, we multiplied the τ values by the standard deviation (also referred to as standard error in our documentation) of the TCC regression trees' predicted values (Equation 2).

Finally, we discovered differences between the LCMS tree mask non-processing areas and the Science TCC non-processing areas because of differences in predictor data used, which was causing TCC data to be eliminated from areas masked by the LCMS non-processing area. To ensure available Science TCC data were included in the NLCD TCC product, we included Science TCC values in the LCMS non-processing areas and used τ percentiles to separate treed from non-treed pixels. The τ percentiles used for Alaska, PRUSVI, and Hawaii were 88, 97, and 95, respectively.

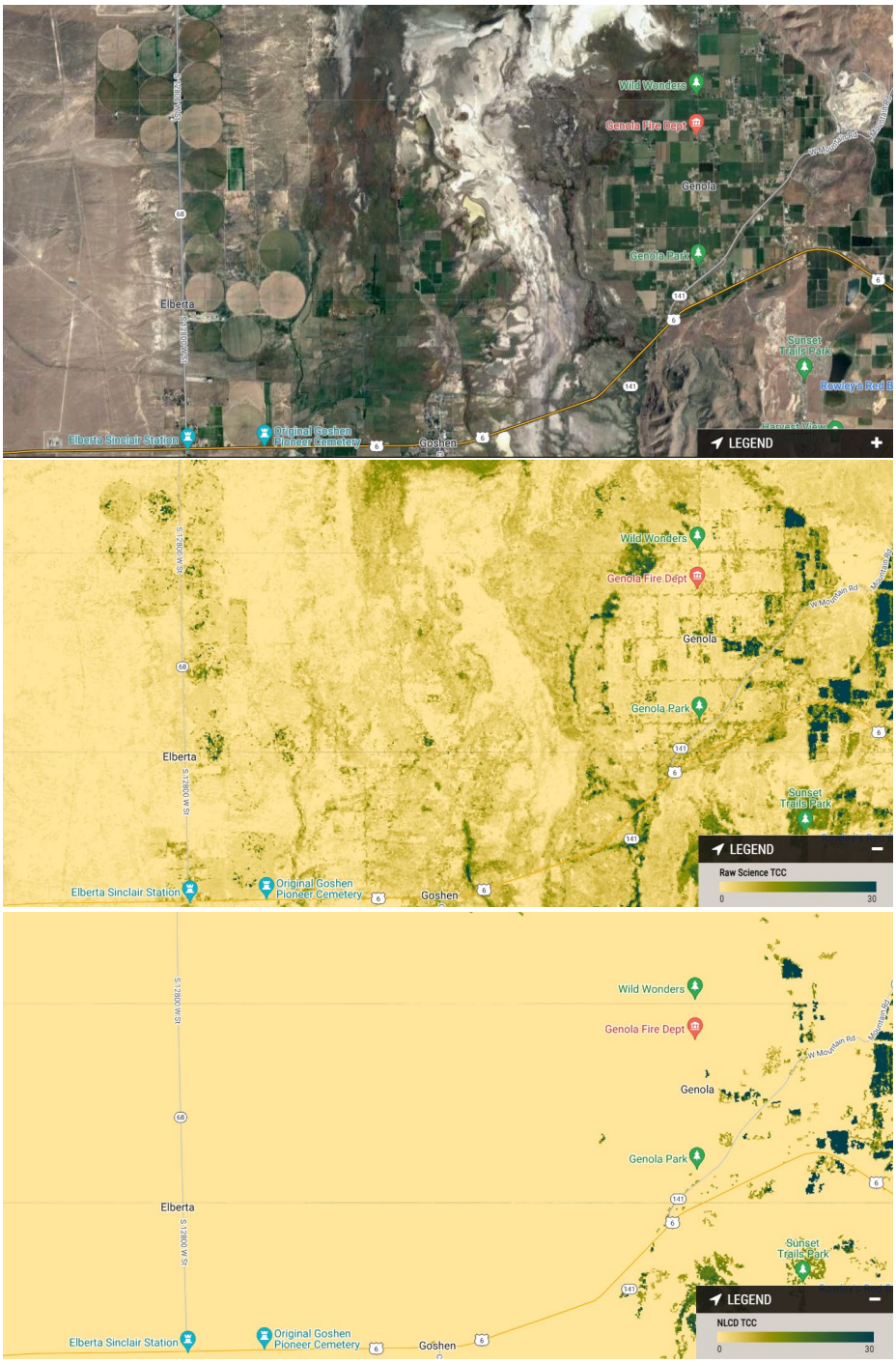

Figure 11. Example area illustrating the results of the non-urban Tree Canopy Cover (TCC) masking. The top image shows high resolution satellite imagery for context. The middle image shows the Science TCC product. The bottom image shows the National Land Cover Database (NLCD) TCC product. Notice how many agriculture and wetland areas had somewhat high TCC values in the raw Science TCC product. These are set to 0 in this filtering step.

Urban TCC Masking

For urban areas, we applied the non-tree masks as were used in the non-urban tree masking to mask the predicted TCC values. However, instead of using the LCMS tree mask, we employed a

different procedure to mask urban predicted TCC values. We used τ, described in the Modeling section, to identify pixels insignificantly different from zero percent TCC. τ simulates a distribution of error values for the model calibration datasets.

For CONUS, when we applied the same τ percentiles throughout urban areas, we observed masking differences between low-density and high-density developed areas (Figure 12). Thus, we separated low-density and high-density developed areas using NLCD developed classes (U.S. Geological Survey 2024). Different τ percentiles adjusted through time were applied to the low-density and high-density areas. For high-density urban areas, the τ statistic at the 90 to 86 percentile was used to threshold the TCC values. For low-density urban areas, the τ statistic at the 88 to 85 percentile was used to threshold the TCC values. These values generally balanced commission and omission of masking non-tree areas in visual exploration of different percentile values.

For OCONUS, we applied the same τ percentiles throughout urban areas. To separate treed from non-treed pixels in OCONUS, we used the 65th τ percentile in SEAK, the 91st τ percentile in PRUSVI, and the 87th τ percentile in Hawaii.

For urban areas, we multiplied the τ values by the standard deviation (also referred to as standard error in our documentation) of the TCC regression trees' predicted values. Urban pixels were assigned to zero percent TCC as is shown in Equation 2.

$$Equation\ 2 \qquad TCC - StDev * \tau \leq 0 \Rightarrow 0\%\ TCC$$

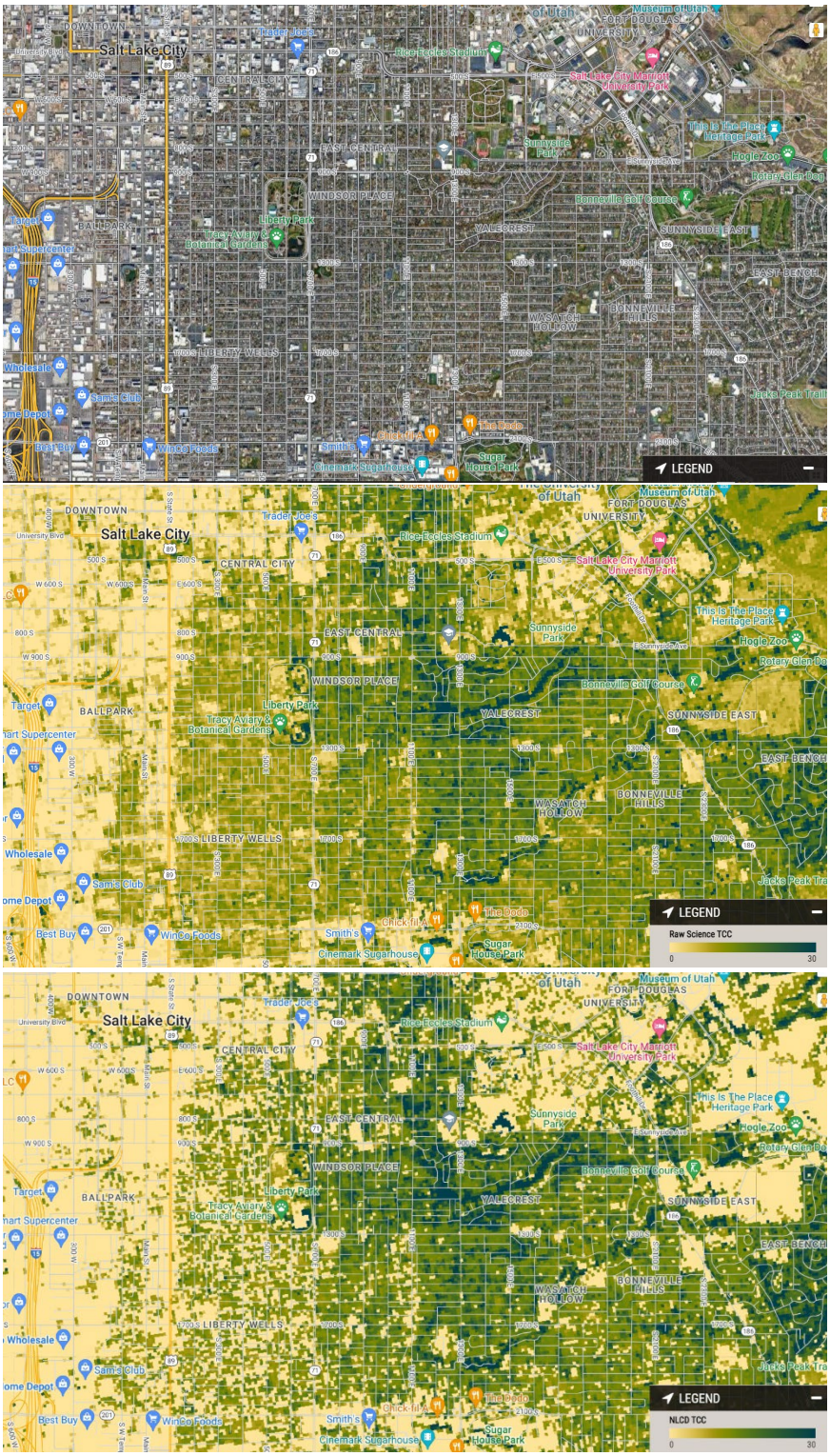

Figure 12. Urban area example from Utah illustrating the impact of the urban low confidence filtering step. The top image shows high resolution satellite imagery for context. The middle image shows the Tree Canopy Cover (TCC) predictions. The bottom image shows the filtered and masked National Land Cover Database (NLCD) TCC product. Notice how many high-density developed areas had somewhat high TCC values in the unfiltered TCC product. The filtering step sets these to zero.

Temporal Filtering

The final post-processing step involved applying methods that reduced interannual noise. First, we removed small interannual changes by only allowing for changes exceeding 10 percent change in TCC each year starting with 1985 or the year with the first available value. This filtering step can have positive and negative impacts on the final outputs, as illustrated in Figure 13, but it simplifies the potential use of NLCD TCC outputs for change detection.

Next, we applied two additional post-processing steps to further reduce interannual noise. Interannual noise persisted in the TCC time series because of the order of operations between the masking and removal of small interannual changes. We applied an algorithm that identified one year positive or negative TCC spikes greater than 10 percent, removed the spikes, and extended the previous year's TCC value through to the spike year. Finally, we updated TCC values to zero when 1) the baseline TCC value was less than 10 percent, 2) the majority of annual TCC predictions were 0, and 3) TCC did not exceed 10 percent following a zero percent TCC prediction in the time series. The result of these post-processing steps is the NLCD TCC v2023.5 product.

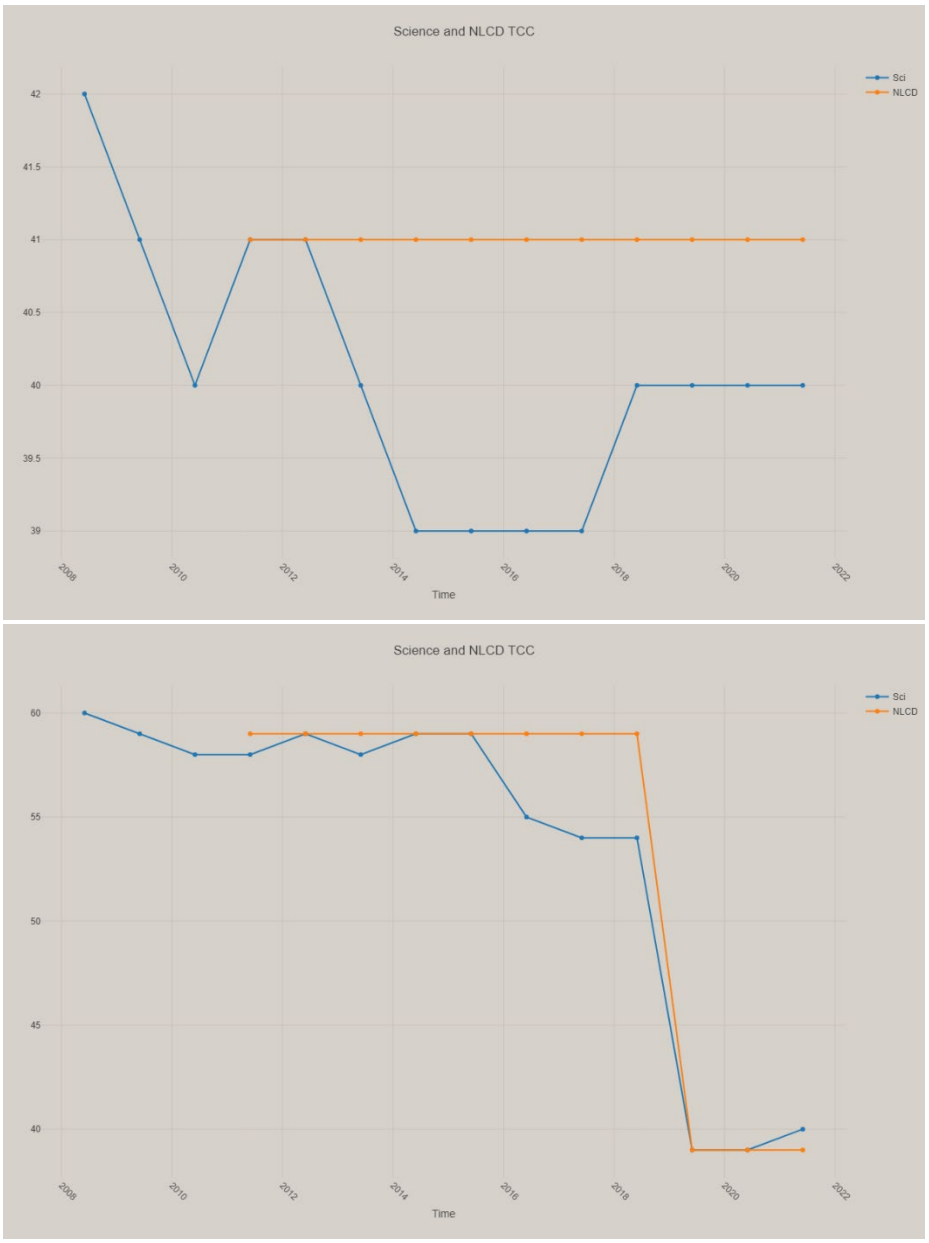

Figure 13. Comparison of the Science Tree Canopy Cover (TCC) and National Land Cover Database (NLCD) TCC outputs illustrating the impact of the temporal filtering routines. This filter is designed to prohibit any interannual change exceeding 10 percent change in TCC in the final NLCD TCC products. The top example illustrates the type of small interannual change common in the Science outputs converted to a steady signal after the filtering routine is applied. The bottom example is selected from a location that experienced a wildfire. In this second example, the filtering routine does result in the elimination of the slight decline in TCC prior to the fire year (2019).

Map Error Assessment

We conducted an independent error assessment over the 2011 NLCD TCC v2023.5 map output. We reserved 30 percent of the reference data to use as map error assessment (see the Reference Data section). Our reference data had varying spatial densities due to 1) different reporting needs from FIA data; 2) different re-measurement frequencies (Reams et al., 2005); and 3) our own filtering process (Table 1). To account for this, we estimated the area weight of each plot by

computing Thiessen polygons for each plot centroid (Figure 14). To avoid extremely large weights for edge plots, any plot on the edge assumed the average of the weight of neighboring non-edge plots' polygon weights.

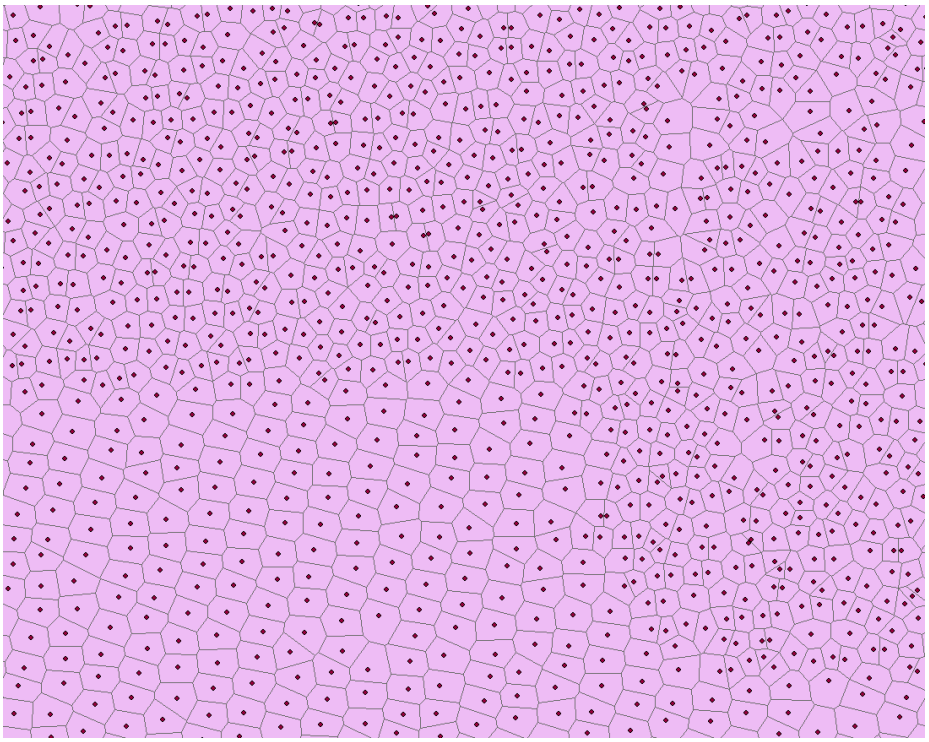

Figure 14. Fuzzed plot locations for an area with their computed Thiessen polygons shown in light gray. The area of each polygon around each plot was then used to weight the plot. This way a plot in a lower density area would receive a higher weight.

We then computed the weighted root mean squared error (RMSE) and mean absolute error (MAE) as shown in Equations 3 and 4.

Equation 3

$$RMSE_{weighted} = \sqrt{\frac{\sum\left(TCC_{predicted} - FIA\_PI\_TCC\right)^2 * Area\_Weights}{\sum Area\_Weights}}$$

Equation 4

$$MAE_{weighted} = \frac{\sum\left(abs\left(TCC_{predicted} - FIA\_PI\_TCC\right)\right) * Area\_Weights}{\sum Area\_Weights}$$

The 2011 CONUS-wide NLCD TCC map weighted RMSE is 13 percent TCC, and the weighted MAE is 8 percent TCC.

Useful Resources

- NLCD TCC Downloads
- Raw Science TCC Downloads
- ESRI Image Services
- TCC Contact Information

References

Bechtold, W. A.; Patterson, P. L. 2005. The enhanced forest inventory and analysis program—National sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station. https://doi.org/10.2737/SRS-GTR-80

Breiman, L. 2001. Random Forests. Machine Learning. 45(1): 5–32. doi: 10.1023/A:1010933404324

Brown, J. F.; Tollerud, H. J.; Barber, C. P.; Zhou, Q.; Dwyer, J. L.; Vogelmann, J. E.; Loveland, T. R.; Woodcock, C. E.; Stehman, S. V.; Zhu, Z.; Pengra, B. W.; Smith, K.; Horton, J. A.; Xian, G.; Auch, R. F.; Sohl, T. L.; Sayler, K. L.; Gallant; A. L.; Zelenak, D.; Reker, R. R.; Rover J. 2020. Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach. Remote Sensing of Environment. 238: 111356. doi: 10.1016/j.rse.2019.111356

Chastain, R.; Housman, I.; Goldstein, J.; Finco, M.; Tenneson, K. 2019. Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM top of atmosphere spectral characteristics over the conterminous United States. Remote Sensing of Environment. 221: 274–285. doi: 10.1016/j.rse.2018.11.012

Coulston, J. W.; Blinn, C. E.; Thomas, V. A.; & Wynne, R. H. 2016. Approximating Prediction Uncertainty for Random Forest Regression Models. Photogrammetric Engineering & Remote Sensing. 82(3): 189–197. doi: 10.14358/PERS.82.3.189

Daubenmire, R. 1959. A canopy-coverage method of vegetational analysis. Northwest Science. 33(1): 43–64.

Flood, N. 2013. Seasonal Composite Landsat TM/ETM+ Images Using the Medoid (a Multi-Dimensional Median). Remote Sensing. 5(12): 6481–6500. doi: 10.3390/rs5126481

Foga, S.; Scaramuzza, P. L.; Guo, S.; Zhu, Z.; Dilley, R. D.; Beckmann, T.; Schmidt, G. L.; Dwyer, J. L.; Hughes, M.J.; Laue, B. 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sensing of Environment. 194: 379–390. doi: 10.1016/j.rse.2017.03.026

Goeking, S. A.; Liknes, G. C.; Lindblom, E.; Chase, J.; Jacobs, D. M.; Benton, R. 2012. A GIS-based tool for estimating tree canopy cover on fixed-radius plots using high-resolution aerial imagery. In R. S. Morin; G. C. Liknes (Eds.), Moving from status to trends: Forest Inventory and Analysis (FIA) symposium 2012. Gen. Tech. Rep. NRS-P-105 (pp. 237–241). Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. https://research.fs.usda.gov/treesearch/42752

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 202: 18–27. doi: 10.1016/j.rse.2017.06.031

Hansen, M. C.; Loveland, T. R. 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sensing of Environment. 122: 66–74. doi: 10.1016/j.rse.2011.08.024

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. Development of a 2001 National Land Cover Database for the United States. Photogrammetric Engineering and Remote Sensing. 70(7): 829–840. doi: 10.14358/PERS.70.7.829

Housman, I.; Campbell, L. S.; Heyer, J. P.; Goetz, W. E.; Finco, M. V.; Pugh, N.; Megown, K. 2022. US Forest Service Landscape Change Monitoring System Methods Version 2021.7. GTAC-10252-RPT3 (No. GTAC-10252-RPT3; pp. 1–24). Salt Lake City, UT: Field Services and Innovation Center-Geospatial Office. doi:10.13140/RG.2.2.19965.23524

Housman, I.; Schleeweis, K.; Heyer, J.; Ruefenacht, B.; Bender, S.; Megown, K.; Goetz, W.; Bogle, S. 2023. National Land Cover Database Tree Canopy Cover Methods v2021.4. GTAC-10268-RPT1: 1–29. Field Services and Innovation Center- Geospatial Office, Salt Lake City, UT.

Kennedy, R. E.; Yang, Z.; Cohen, W. B. 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sensing of Environment. 114(12): 2897–2910. doi: 10.1016/j.rse.2010.07.008

Kennedy, R. E.; Yang, Z.; Gorelick, N.; Braaten, J.; Cavalcante, L.; Cohen, W. B.; Healey, S. 2018. Implementation of the LandTrendr Algorithm on Google Earth Engine. Remote Sensing. 10(5): 691. doi: 10.3390/rs10050691

Lin, L.; Di, L.; Zhang, C.; Guo, L.; Di, Y.; Li, H.; Yang, A. 2022. Validation and refinement of cropland data layer using a spatial-temporal decision tree algorithm. Scientific Data. 9(1): 63. doi: 10.1038/s41597-022-01169-w

Moisen, G.G.; Meyer, M.C.; Schroeder, T.A.; Liao, X.; Schleeweis, K.G.; Freeman, E.A.; Toney, C. 2016. Shape selection in Landsat time series: A tool for monitoring forest dynamics. Global Change Biology. 22(10): 3518–3528. doi: 10.1111/gcb.13358

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12(85): 2825–2830. doi: https://doi.org/10.48550/arXiv.1201.0490

Potapov, P.; Hansen, M. C.; Pickens, A.; Hernandez-Serna, A.; Tyukavina, A.; Turubanova, S.; Zalles, V.; Li, X.; Khan, A.; Stolle, F.; Harris, N.; Song, X.-P.; Baggett, A.; Kommareddy, I.; Kommareddy, A. 2022. The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. Frontiers in Remote Sensing. 3: 856903. doi: 10.3389/frsen.2022.856903

Powell, S. L.; Cohen, W. B.; Healey, S. P.; Kennedy, R. E.; Moisen, G.G.; Pierce, K. B.; Ohmann, J. L. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. Remote Sensing of Environment. 114(5): 1053–1068. doi: 10.1016/j.rse.2009.12.018

R Core Team. 2020. R: A language and environment for statistical computing [R]. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Reams, G. A.; Smith, W. D.; Hansen, M. H.; Bechtold, W. A.; Roesch, F. A.; Moisen, G. G. 2005. The forest inventory and analysis sampling frame. Gen. Tech. Rep. SRS-80 (No. Gen. Tech. Rep. SRS-

80; pp. 21–36). Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. Retrieved from U.S. Department of Agriculture, Forest Service, Southern Research Station website: https://research.fs.usda.gov/treesearch/20376

U.S. Census Bureau, Geography Division. 2024. TIGER/Line Shapefiles, 2024: Urban Areas [Shapefile]. Retrieved from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

U.S. Geological Survey. 2019. USGS 3D Elevation Program Digital Elevation Model [Data set]. https://developers.google.com/earth-engine/datasets/catalog/USGS_3DEP_10m. https://developers.google.com/earth-engine/datasets/catalog/USGS_3DEP_10m

U.S. Geological Survey. 2024. Annual NLCD Collection 1 Science Products: U.S. Geological Survey data release [Data set]. Sioux Falls, SD: U.S. Geological Survey. doi: https://doi.org/10.5066/P94UXNTS

U.S. Department of Agriculture, Forest Service. 2025. USFS Landscape Change Monitoring System Conterminous United States version 2024-10 (Version 2024-10) [Data set]. https://data.fs.usda.gov/geodata/rastergateway/LCMS/index.php

U.S. Department of Agriculture, National Agricultural Statistics Service. 2008. Published crop-specific data layer [Data set]. https://nassgeodata.gmu.edu/CropScape USDA-NASS. https://nassgeodata.gmu.edu/CropScape USDA-NASS

Vogeler, J. C.; Braaten, J. D.; Slesak, R. A.; Falkowski, M. J. 2018. Extracting the full value of the Landsat archive: Inter-sensor harmonization for the mapping of Minnesota forest canopy cover (1973–2015). Remote Sensing of Environment. 209: 363–374. doi: 10.1016/j.rse.2018.02.046

Wickham, J.; Homer, C.; Vogelmann, J.; McKerrow, A.; Mueller, R.; Herold, N.; Coulston, J. 2014. The Multi-Resolution Land Characteristics (MRLC) Consortium—20 Years of Development and Integration of USA National Land Cover Data. Remote Sensing. 6(8): 7424–7441. doi: 10.3390/rs6087424

Zhu, Z.; Woodcock, C. E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment. 118: 83–94. doi: 10.1016/j.rse.2011.10.028

Author contributions

Ian Housman – Conceptualization, methodology, software, validation, formal analysis, investigation, writing – original draft, visualization

Joshua Heyer – Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing

Bonnie Ruefenacht – Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing

Karen Schleeweis – Conceptualization, methodology, validation, formal analysis, writing

Kevin Megown – Conceptualization, resources, funding acquisition, supervision

Seth Bogle – Project administration

Jaclyn Reischmann – Visualization, project administration,

Daniel Ryerson – Project administration, funding acquisition