



National Land Cover Database Tree Canopy Cover Methods

Version: 2021.4

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

Geospatial Technology and Applications Center (GTAC)

125 S. State Street, Suite 7105, Salt Lake City, Utah 84138

apps.fs.usda.gov/gtac/

Ian Housman

Lead Technical Consultant
Senior Remote Sensing Specialist
RedCastle Resources, onsite contractor
ian.housman@usda.gov

Karen Schleeweis

Science Team Lead
Rocky Mountain Research Station, FIA
karen.schleeweis@usda.gov

Josh Heyer

Production Lead
Geospatial Specialist
RedCastle Resources, onsite contractor
joshua.heyser@usda.gov

Bonnie Ruefenacht

Production Lead
Geospatial Programmer
RedCastle Resources, onsite contractor
bonnie.ruefenacht@usda.gov

Stacie Bender

Physical Science Information Specialist
Resource Mapping, Inventory, and Monitoring (RMIM)
stacie.bender@usda.gov

Kevin Megown

Program Leader
Resource Mapping, Inventory, and Monitoring (RMIM)
kevin.megown@usda.gov

Wendy Goetz

Project Contributor
RedCastle Resources, onsite contractor
wendy.goetz@usda.gov

Seth Bogle

Geospatial Project Manager
RedCastle Resources, onsite contractor
michael.bogle@usda.gov

USDA Non-Discrimination Statement

In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, gender identity (including gender expression), sexual orientation, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotope, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at [How to File a Program Discrimination Complaint](#) and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by: (1) mail: U.S. Department of Agriculture, Office of the Assistant Secretary for Civil Rights, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410; (2) fax: (202) 690-7442; or (3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

Contents

Executive Summary	1
Background	1
Methods	2
STUDY AREAS	2
COMPUTING PLATFORMS	3
REFERENCE DATA	3
REFERENCE DATA COLLECTION	3
REFERENCE DATA FILTERING	4
MODEL PREDICTOR DATA	6
REMOTE SENSING SPECTRAL DATA	6
TERRAIN DATA	13
CROPLAND DATA	13
SUMMARY	13
MODELING	14
SCIENCE TCC PRODUCT MODELING	14
SCIENCE PRODUCTS	16
NLCD TCC POST-PROCESSING METHODS	16
SEPARATING URBAN AND NON-URBAN AREAS	17
NON-URBAN NON-TREE TCC MASKING	17
URBAN NON-TREE TCC MASKING	19
TEMPORAL FILTERING	21
MAP ERROR ASSESSMENT	22
Useful Resources	23
References	24

Executive Summary

The National Land Cover Database (NLCD) Tree Canopy Cover (TCC) is a remote sensing-based map output produced by the United States Department of Agriculture, Forest Service (Forest Service). The 2021 NLCD TCC release includes annual maps from 2011-2021 across the Conterminous United States (CONUS), coastal Alaska, Puerto Rico and the US Virgin Islands (PRUSVI), and Hawaii. Science TCC products are also available annually from 2008 to 2021. 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 2021.4. A forthcoming manuscript will provide a more comprehensive summary of TCC 2021.4 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 this version of the production methods was to enable more frequent time steps, a longer temporal extent, and more coherent temporal signals in the TCC products. This goal was achieved by using GEE storage and compute capacity for image processing, and by incorporating time-series-based approaches already utilized by Forest Service programs like 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 on the [Multi-Resolution Land Characteristics \(MRLC\) consortium website](#).

Annual Science TCC products are available for download on the [FSGeodata Clearinghouse website](#).

Background

The Forest Service produces the NLCD TCC product as a member of the 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 and 2016 releases and the most recent 2021 release (Coulston et al., 2012; Yang et al., 2018). 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.

All production methods are developed as a joint effort between the NLCD TCC Science Team under the Forest Inventory and Analysis (FIA) program and the NLCD TCC Production Team at the Geospatial Technology and Applications Center (GTAC). Methods are revisited and potentially changed with each new release. To allow for easy tracking of which product version you are using, all products and documentation will share the following naming convention: YYYY.v, where “YYYY” denotes the most recent year mapped, and the “v” denotes the version of the production methods used. Version 2021.4 is the first to employ this naming convention.

Annual nationwide mapping programs are becoming more common (Brown et al., 2020; Housman et al., 2022). One goal of the TCC version 2021.4 release was to leverage Forest Service LCMS methods where appropriate. As a result, NLCD TCC now shares much of the same image inputs and image pre-processing workflows used by the Forest Service LCMS. While annual NLCD TCC products are not yet being released by MRLC, annual NLCD TCC and Science TCC products (raw direct model outputs and per-pixel model standard deviation – also referred to as “standard error”) extending from 2008-2021 are being released by the Forest Service.

Since NLCD TCC 2021.4 shares some methods with the 2016 NLCD TCC release and LCMS, this document will pull from the existing 2016 NLCD TCC methods documentation (Ruefenacht et al., 2022) as well as the LCMS methods documentation (Housman et al., 2022) when appropriate.

NLCD TCC products cover the Conterminous United States (CONUS), and outside-CONUS (OCONUS) mapping areas Coastal Alaska, Puerto Rico and US Virgin Islands (PRUSVI), and Hawaii. CONUS products will be released spring 2023, with the other study areas following in summer/fall 2023. The NLCD and Science TCC product suite includes 30 m spatial resolution maps of percent tree canopy cover from 2008-2021.

Methods

Study Areas

The 2021 TCC product suite extends across CONUS, coastal Alaska, Hawaii, and PRUSVI (figure 1).

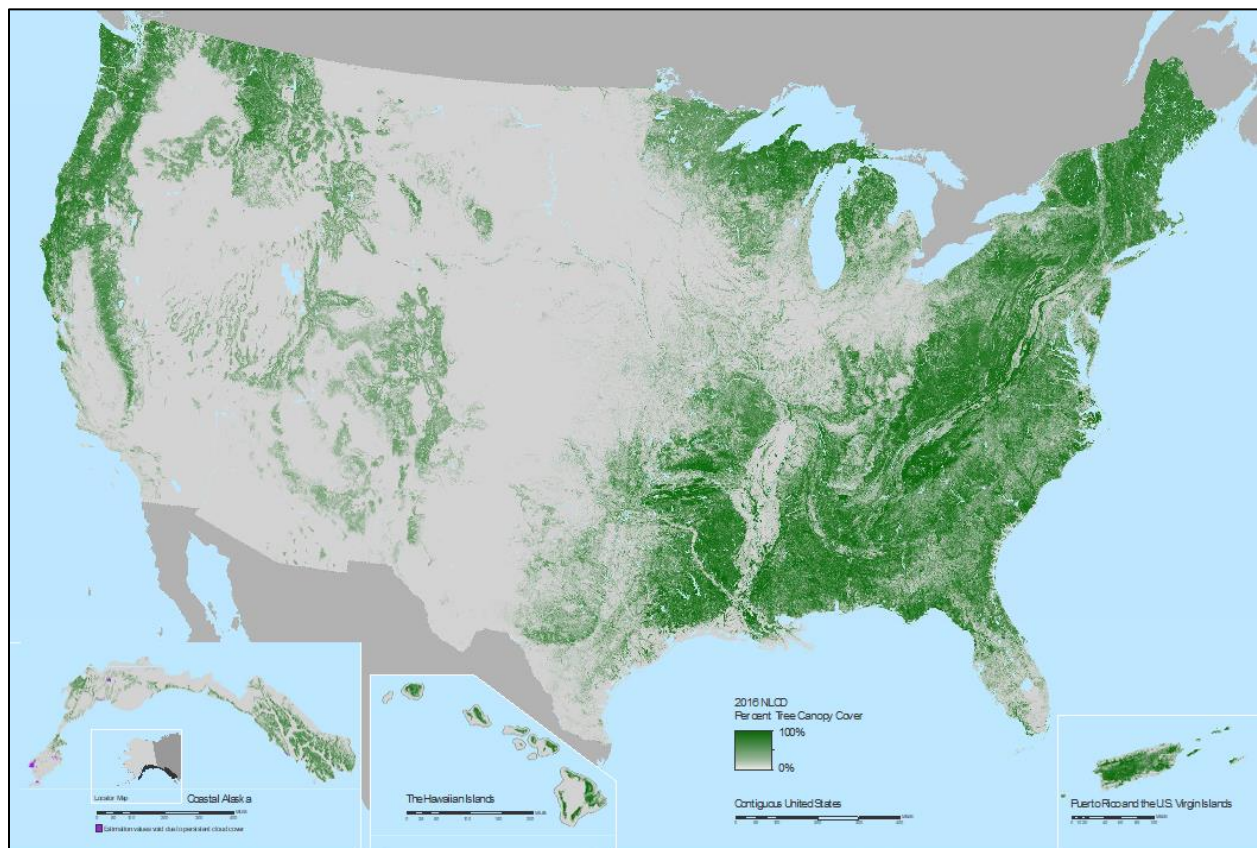


Figure 1.—Depiction of the tree canopy cover study areas for CONUS, coastal Alaska, Hawaii, and Puerto Rico and the US Virgin Islands (PRUSVI).

All study areas except coastal Alaska match those of NLCD land cover and urban impervious cover products, produced by USGS. Specifically, CONUS consists of the 48 contiguous states and some surrounding water. 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, while the US Virgin Islands portion of PRUSVI consists of St. Croix, St. Thomas, and St. John. Coastal Alaska includes a 219,000 km² region of southeast and south-central coastal Alaska (figure 1).

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, along with common data processing methods and computing infrastructure to process these data.

To maintain confidentiality of the reference data locations collected over the FIA sample grid (Bechtold and Patterson 2005), we use R (R Core Team, 2010) and the Python package [Scikit-Learn](#) (Pedregosa et al., 2011) on local compute resources at GTAC to extract predictor data over reference locations, build the 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 (Goeking et al., 2012; figure 2). To compute a plot's estimated tree canopy cover percentage, a grid of 109 dots (with a spacing of 26.25 feet) are placed within a 144-foot radius circle that is centered on the plot. These 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 assigned a confidence level in their photo interpretations for each plot. The plot locations used for reference data collection were located on the FIA grid (Reams et al., 2005).

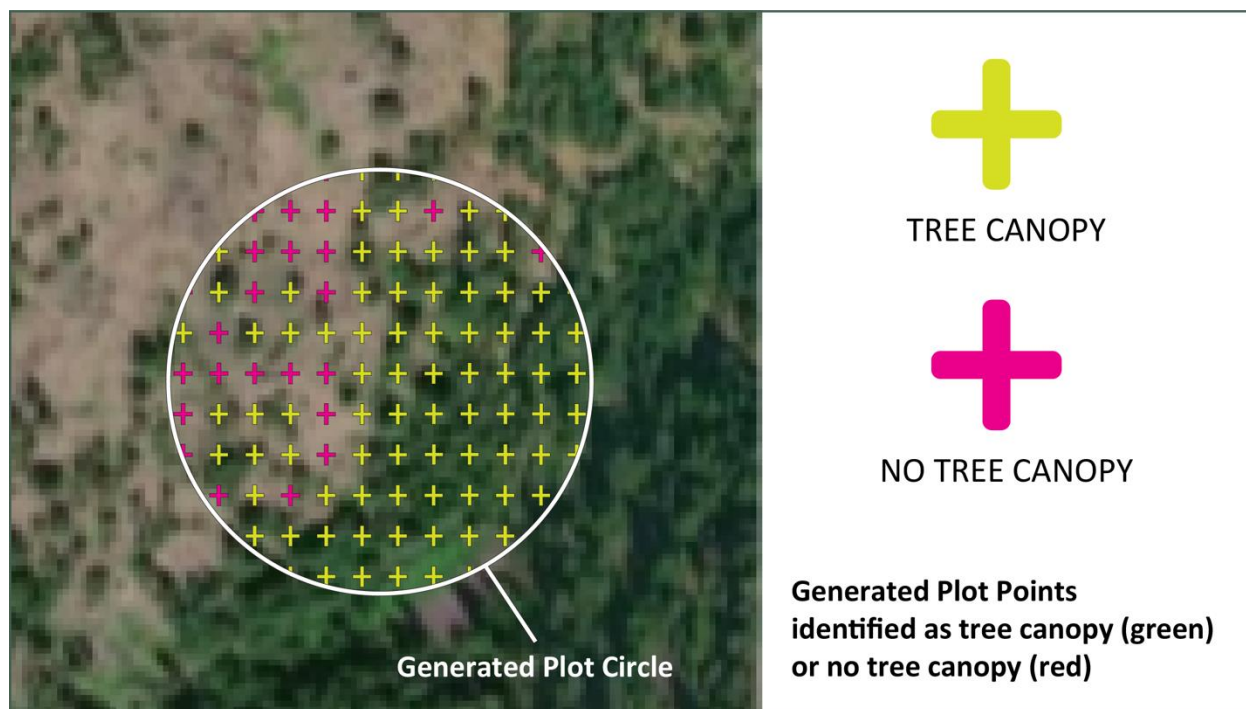


Figure 2. Depiction of the Canopy Cover tool used to develop tree canopy cover reference data.

For this project, tree canopy is defined as:

- Tree – live tally trees, saplings and seedlings – where tally trees are all species listed in the [FIA Master Tree Species List](#). 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 (Daubenmire, 1959), and ignoring overlap among multiple layers of a species.

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 (defined above).

Reference Data Filtering

To minimize error in the CONUS and OCONUS reference data, we omitted any plot for the following reasons:

- Low interpreter confidence
- Spectrally unstable (possible change)
- LCMS Change (Loss or Gain)
- Plot fell over an ocean waterbody
- No cloud or cloud shadow-free spectral information available
- Reference TCC > 0 where manual image interpretation indicated TCC = 0

Of the 63,010 FIA CONUS plots, the final number of plots included as reference data was 55,356 (table 1). These plots were then randomly divided into model calibration (70%) and map error assessment (30%) groups.

For coastal Alaska, tree canopy cover data was photo-interpreted over 2,035 FIA plots. Plots over offshore water or that received a low confidence score by the photo-interpreter₅ were discarded. The final number of plots used as reference data were 1,397 (table 1). These plots were randomly divided into model calibration (80%) and map error assessment (20%) groups.

For Hawaii, tree canopy cover was photo-interpreted over 1,284 FIA plots. For map error assessment, a random set of 325 non-FIA plots were photo-interpreted.

For PRUSVI, 1,965 FIA plots were available. We omitted plots for the following reasons:

- Outside of PRUSVI NLCD study area
- Experimental plots
- Low interpreter confidence
- Reference TCC = 0 where manual image interpretation indicated TCC > 0

To increase the total number of reference points in PRUSVI, we intensified the FIA grid to add 381 points. The final number of plots used as reference data were 980 (table 1). For map error assessment, a random set of 324 non-FIA plots were photo-interpreted.

Table 1. FIA plots were removed from reference data due to low photo-interpretation quality scores, duplicate or experimental plots, locations over oceans or unavailable data, unstable spectral data, inaccurate zero canopy photo-interpretation (PI), and PI that spectrally changed according to LCMS change data between 2009-2012 (within +/- 1 year of NAIP PI acquisition window). The following tables show the number of plots removed and the final number of plots used as reference data.

CONUS		
FIA PI reason for removal	Number of PI removed	PI 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	
<i>Final number of PI plots used</i>		55,356

Coastal Alaska		
FIA PI reason for removal	Number of PI removed	PI Count
All FIA Plots		2,035
Low photo-interpretation quality scores	184	
Ocean or unavailable spectral data	454	
<i>Final number of PI plots used</i>		1,397

PRUSVI		
FIA PI reason for removal	Number of PI removed	PI Count
All FIA Plots		1,965
Outside PRUSVI	1,233	
Experimental plots	126	
Low photo-interpretation quality scores	1	

Disagreement with zero canopy cover and imagery	6	
Intensified plots added		381
Final number of PI plots used		980

Hawaii		
FIA PI reason for removal	Number of PI removed	PI Count
All FIA Plots		1,709
Non-FIA Plots	325	
Low photo-interpretation quality scores	100	
Final number of PI plots used		1,284

Model Predictor Data

Our model predictor data consists of Landsat and Sentinel-2 imagery fitted by the LandTrendr temporal segmentation algorithm, topographic information from the USGS 3-Dimensional Elevation Product (3DEP), and the USDA Cropland Data Layer. Descriptions of each of these datasets are provided below.

Remote Sensing Spectral Data

Data Preparation

We use United States Geological Survey (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 2022. We do not use surface reflectance data because the Sentinel-2 surface reflectance data available within GEE are terrain-corrected, which makes it difficult to use with Landsat surface reflectance data that are not terrain-corrected.

For cloud masking Landsat data, we apply the CFmask cloud and cloud shadow masking algorithm (Foga et al., 2017), which is an implementation of Fmask 2.0 (Zhu and Woodcock, 2012), as well as the cloudScore algorithm (Chastain et al., 2019). For cloud masking Sentinel-2 data, we utilize the s2Cloudless algorithm (Zupanc, 2017) outputs available through GEE. We mask 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 can be accessed in the GTAC GEE data processing and visualization library:

- [GeeViz Module on PyPI](#)
- [GeeViz Module on GitHub](#)

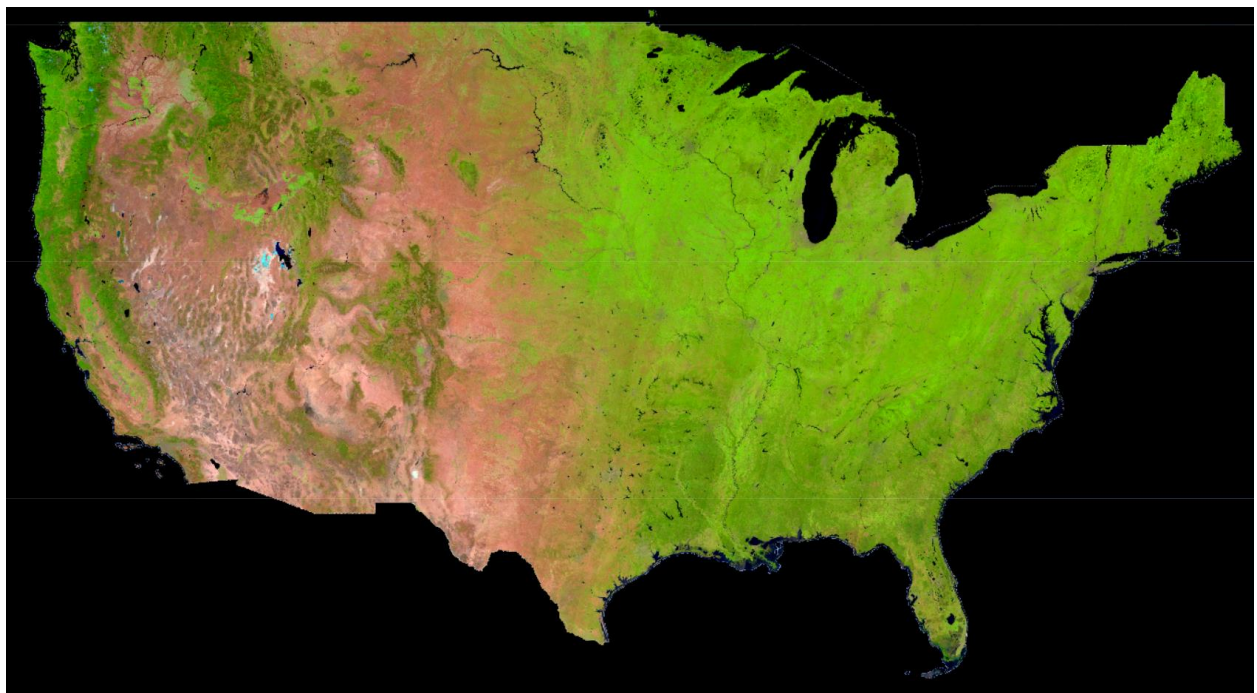
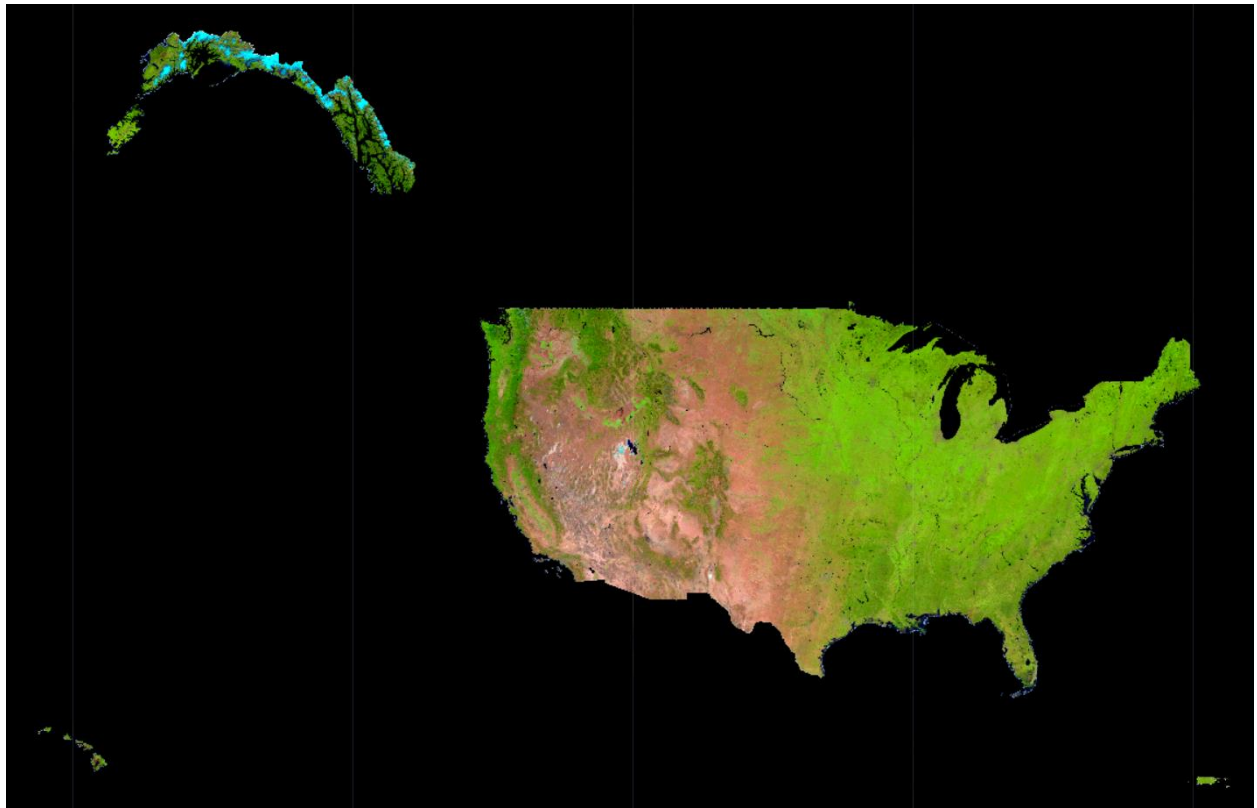
Annual Compositing

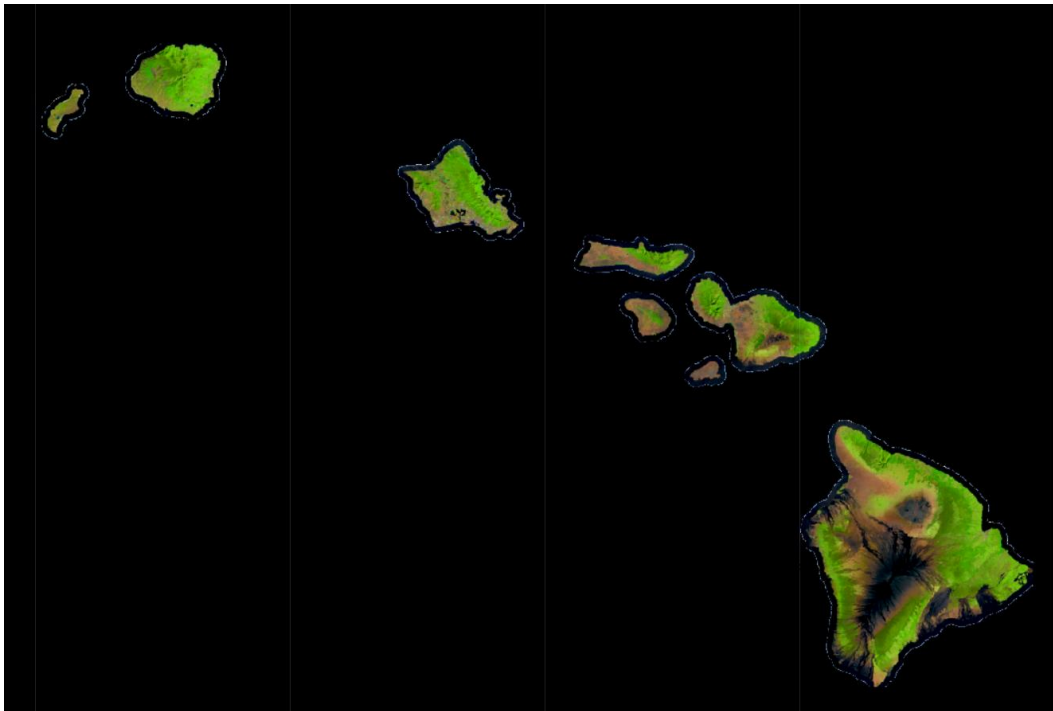
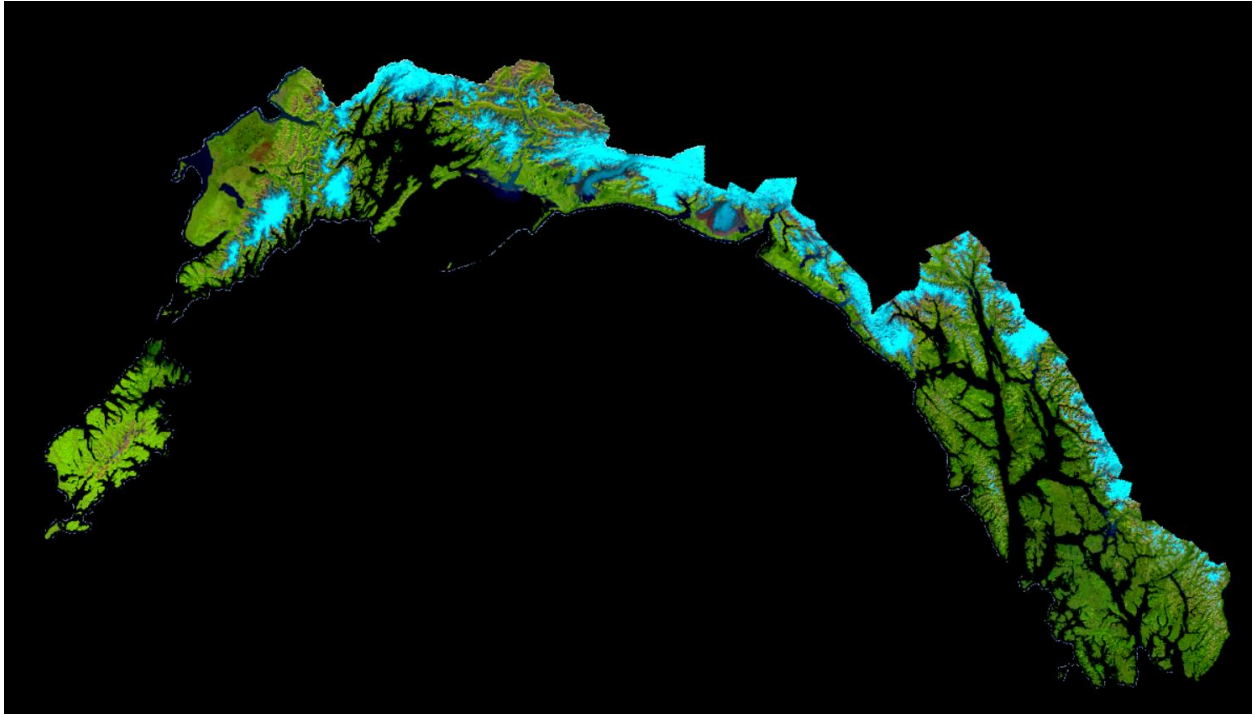
We utilize annual composites of the masked Landsat and Sentinel-2 data as inputs for LandTrendr. Annual composite values are the geometric medoid of all values not masked as cloud or cloud shadow from a specified date range for each year. Due to differences in data availability and seasonality, we adapt 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
CONUS	June 1	September 30	July 1	September 1
Coastal AK	June 15	September 15	June 15	September 15
PRUSVI	June 1	May 31	June 1	May 31
HI	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 et al., 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 include in the feature space are green, red, near infrared (NIR), first shortwave infrared (SWIR1), and second shortwave infrared (SWIR2). We omit the blue band because it is more prone to atmospheric scattering and can inappropriately influence the medoid algorithm. Any pixel that does not have a cloud or cloud shadow free value for a given year is left as null and excluded from the composite for that year. The 2020 composite images for CONUS and coastal Alaska are shown in figure 3 as an example.





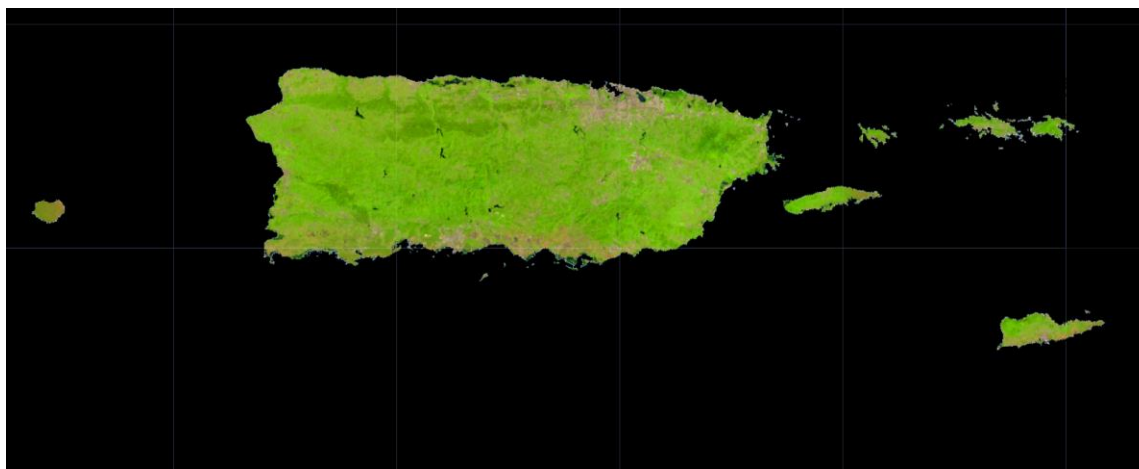


Figure 3. Example of the 2020 composites used in NLCD TCC. The red, green, and blue channels used in these composites are the second shortwave infrared (SWIR2), near infrared (NIR), and red bands, respectively. The top image shows all study areas. The following images show CONUS, coastal Alaska, Hawaii, and PRUSVI study areas from top to bottom respectively.

With the workflow outlined above, two separate sets of annual composite series from 1984 to 2022 are 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 a stripe 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), 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 use 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 iteratively breaks the time series of annual composites and returns a set of segments. Each segment has a start and end year, and a start and end fitted value at the start and end vertices respectively (figure 4). From this information, we assign each band/index for each year the fitted value.

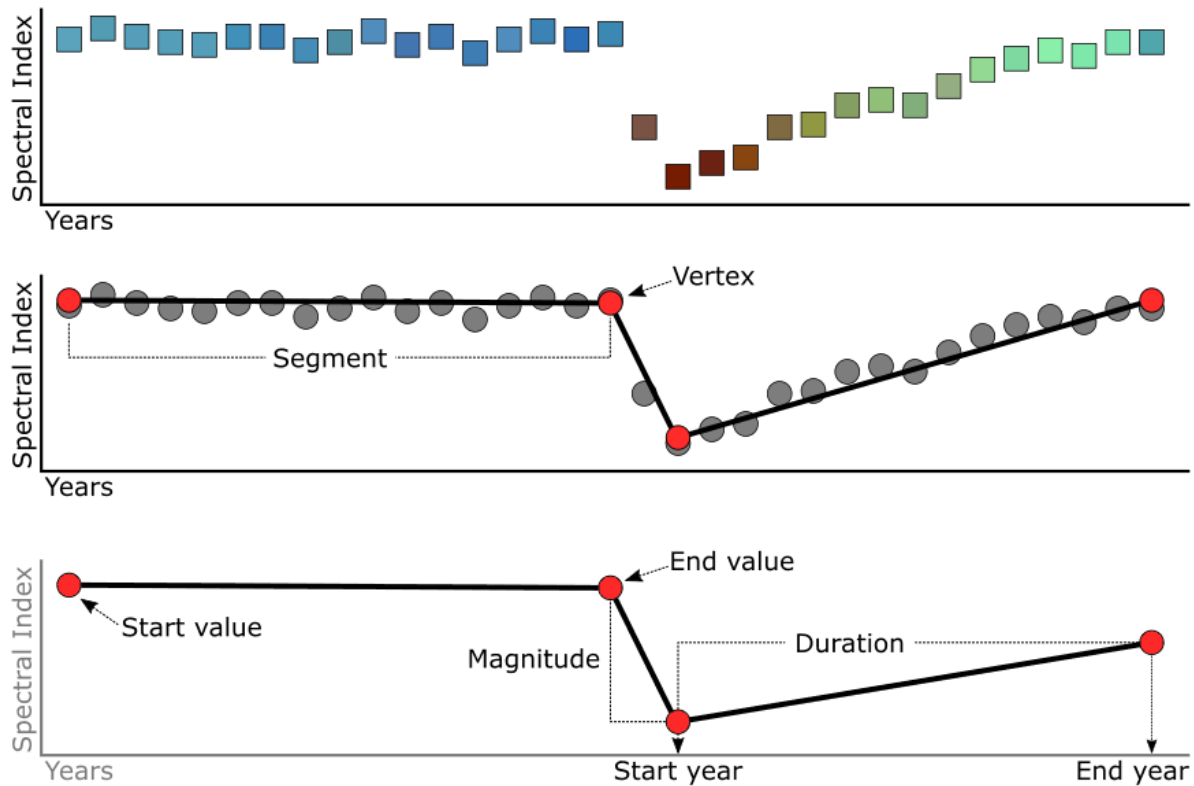


Figure 4. Illustration from the [LandTrendr Google Earth Engine \(GEE\) Guide](#) depicting how LandTrendr breaks a time series and the information that can be taken from the output.

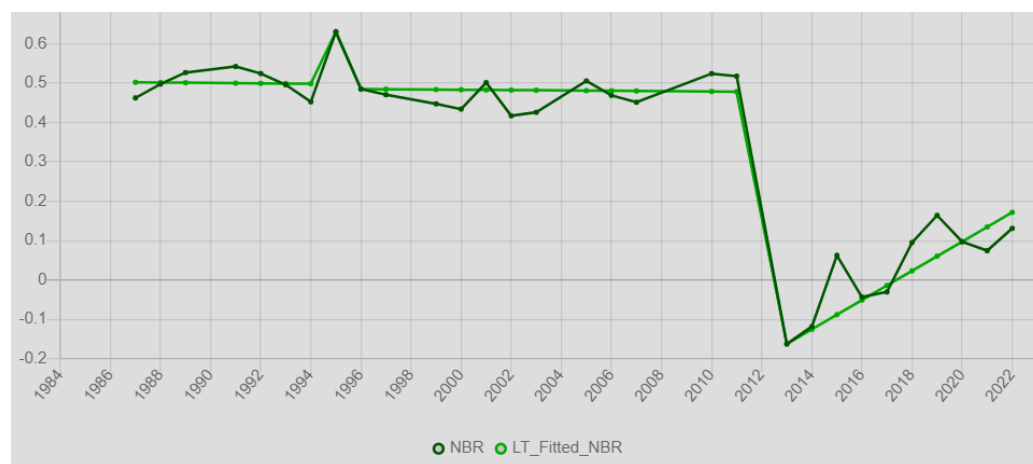
We use the GEE version of LandTrendr outlined in Kennedy et al. (2018). The parameters that are used are the same as those in Kennedy et al. (2018) (table 3). Further documentation of the LandTrendr method used can be found in the [GEE documentation](#).

Table 3. LandTrendr parameters we used in temporal segmentation.

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.

Summary

Figure 5 illustrates how one of our predictor variables (the normalized burn ratio (NBR)) is segmented by LandTrendr for two example pixels. The pixel depicted in the top chart shows a fire event, while the bottom chart 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 use in our models. Only the 2011 LandTrendr values were used for model calibration, while all respective years were used for model application.



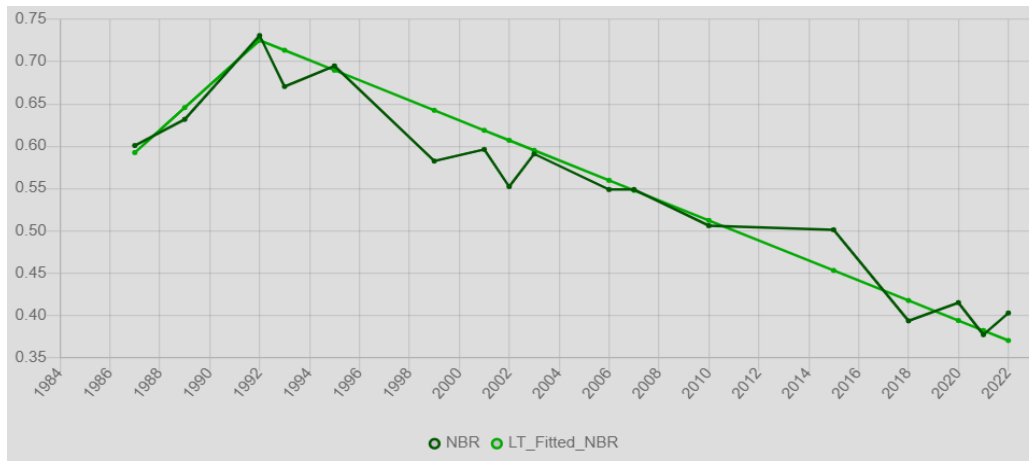


Figure 5. An example of raw medoid composite and LandTrendr fitted NBR values for a single pixel. The pixel depicted in the top chart shows a fire event, while the bottom chart shows insect-related tree mortality.

Terrain Data

We use terrain metrics to provide elevation, slope, and aspect, information to the model. The specific variables used are:

- Elevation
- Aspect
- Sine (Aspect)
- Cosine (Aspect)
- Slope

For all study areas, the 10 m USGS 3D Elevation Program (U.S. Geological Survey, 2019) was used. All resampling from 10 m to 30 m spatial resolution was performed using cubic convolution.

Cropland Data

For CONUS, we also use the USDA National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) cropland layers to help inform the model on locations of agricultural areas (Lin et al., 2022). For each CONUS model year (2008-2021), binary agriculture data are produced by classifying all non-tree CDL crops as agriculture, and everything else non-agriculture. No cropland layers were used for the OCONUS.

Summary

All variables covered in this section are utilized in the methods outlined below. Table 4 shows a full list of all predictor variables considered for modeling.

Table 4. List of NLCD TCC model predictor variables. This table depicts both annual values, which are different for each year of the analysis period, and static terrain values, which remain constant. Sets 1 and 2 both include the following sensors: Landsat 4, Landsat 5, Landsat 7, Landsat 8, Landsat 9, Sentinel 2a and Sentinel 2b.

Predictor Variables	Set 1: (No Landsat 7 after 2002)	Set 2: (Includes Landsat 7 until 2015)
Spectral Bands		
blue		
green		
red		
NIR		
SWIR1		
SWIR2		
Indices/Transformations		
NDVI		
NBR		
NDMI		
NDSI		
brightness		
greenness		
wetness		
brightness/ greenness angle		
Ag		
Cropland Data Layer (CDL)	(CONUS only)	NA
Terrain (static)		
Elevation		NA
Slope		NA
cosine(Aspect)		NA
sine(Aspect)		NA

Modeling

Science TCC Product Modeling

NLCD TCC utilizes the random forest modeling method (Breiman, 2001). Random forest randomly selects a subset of the predictor variables and training sites in many different regression trees. Each of the trees predicts a TCC value, the mean of which is used as the final modeled TCC value.

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

Due to its size and wide variety of ecotones, CONUS modeling was broken up into fifty-four 480x480 km tiles. For each tile, a unique random forest model was built using 2011 fitted LandTrendr, 2011 CDL, and

terrain data. All reference data that were part of the 70 percent available for model calibration that intersected tiles within a 5x5 window around the center tile were used to train the random forest model. That model was then applied to the center tile. Figure 6 illustrates how the moving window approach covered CONUS. A moving window approach was not used for the OCONUS. For each OCONUS mapping area one random forest model was trained and applied.

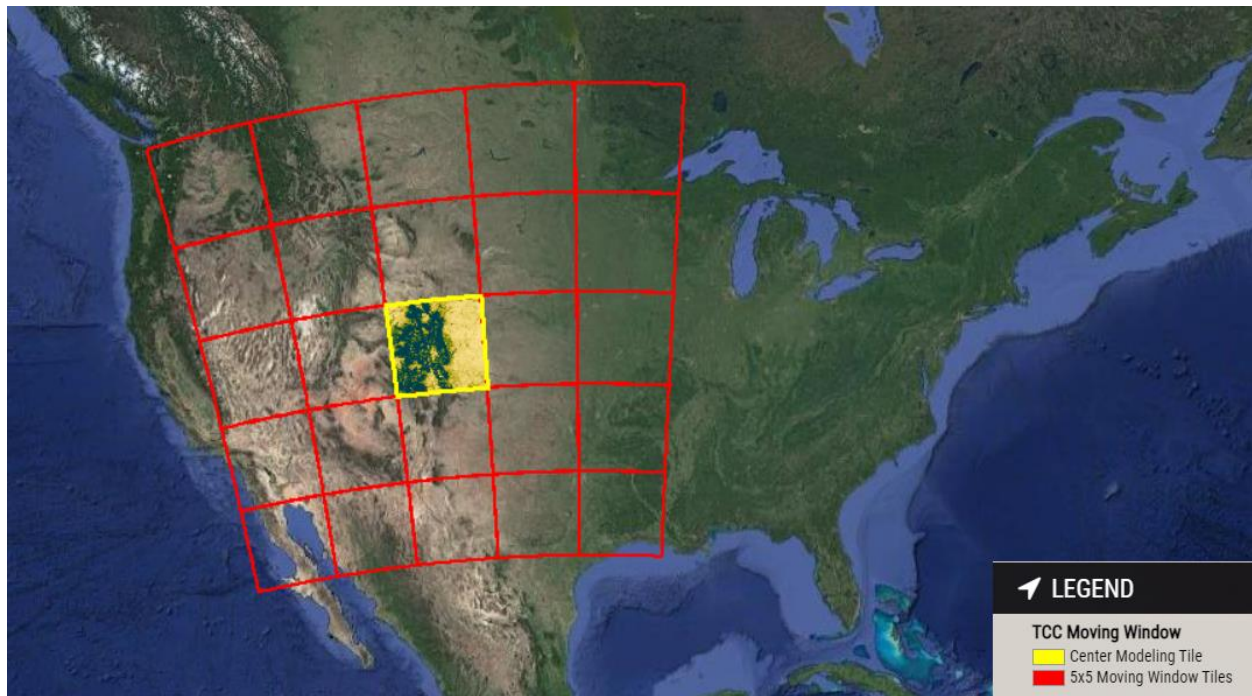


Figure 6. Illustration of 5x5 480x480 km moving window modeling method. This frame illustrates the adjacent tiles in red where training data were pulled from, and the yellow tile in the center the model was then applied to. The raw Science 2011 TCC map is depicted as it is predicted for this tile. Note this animation 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 (Powell et al., 2010; Hansen and Loveland, 2012; Moisen et al., 2016; Vogeler et al., 2018; Potapov et al., 2022). Following this logic, we applied the 2011 model to predict TCC from 2008 to 2021. The last year of the image time series fitted by LandTrendr (2022) 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 we simulated a model error statistic referred to as tau, following guidance from Coulston et al. (2016). For CONUS, tau percentiles were simulated for each of the 54 tiles, and for OCONUS tau percentiles were simulated for the mapping area. The intent of tau 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, in which 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 tau 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 the regression trees within the random forest model predicted (figure 7), along with the tau tables for each of the 54CONUS models and for each OCONUS model. These outputs are available annually from 2008-2021 through the [Forest Service TCC distribution website](#).

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

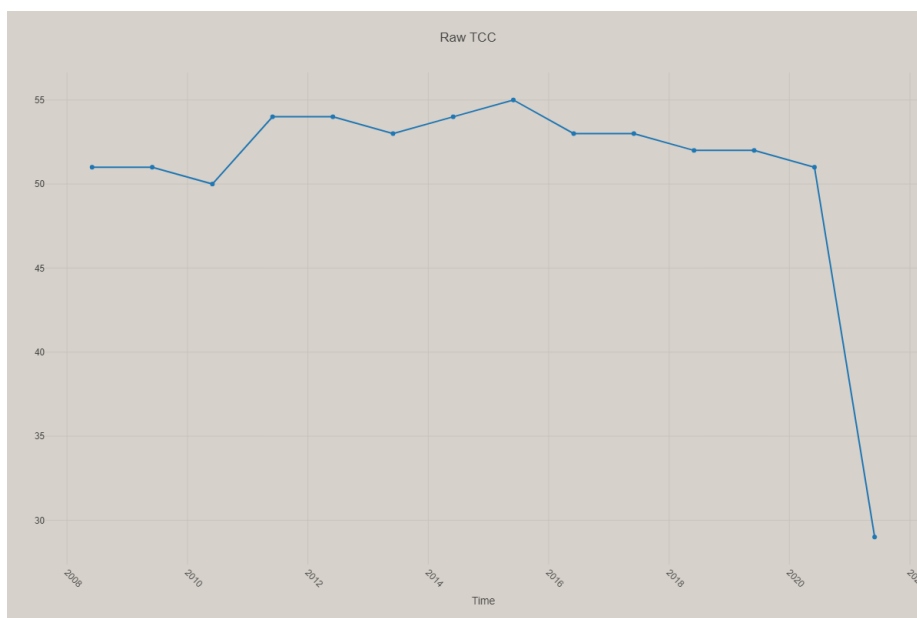


Figure 7. Example of a pixel of the raw Science TCC output. Notice values will bounce around a little each year. This pixel was initially a healthy forest, followed by a long period of gradual decline of TCC from 2015-2020. Then in late summer 2020 (2021 mapped year), this area experienced a wildfire.

NLCD TCC Post-Processing Methods

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

- For non-urban areas:
 - Mask implausible TCC values that should be 0% TCC
 - Mask small clumps of trees
- For urban areas:

- Mask areas with low TCC and high model uncertainty and simulated error in urban areas
- For all mapping areas, filter out small inter-annual fluctuations that could confound interpretations of change common in the time series of predicted TCC values.

Separating Urban and Non-Urban Areas

Our first goal was to mask areas with predicted TCC values greater than 0 that were likely 0. To mitigate masking highly fragmented tree canopy cover common in urban settings, we separated urban and non-urban areas. We used 2018 TIGER U.S. Census Block data, and the LCMS land use developed class as our urban mask.

Non-Urban Non-Tree TCC Masking

For all non-urban areas, we first created a mask annually to mitigate errors related to false positive (TCC>0%), and salt and pepper errors from tiny groups of trees or single trees in the landscape, which are harder to characterize accurately. For a given year's TCC output, we used the following mask primitives:

- 3 year moving window of LCMS non-tree class mode
 - Any pixel where the mode of a 3-year window (+/- 1 year from year of interest) was any of the non-tree LCMS classes (any class without a tree as a primary or secondary sub-class).
- Crop mask
 - NASS CDL non-tree crop areas for the given year of interest for CONUS.
 - No data were used to identify non-tree crop in the OCONUS.
- Water mask
 - NLCD (2019 release) land cover water data from the closest year available to the year of interest for CONUS.
 - For coastal Alaska and PRUSVI the annual LCMS Landcover product water class.
 - For Hawaii the JRC Yearly Water Classification History, v1.4 (Jean-Francois Pekel et al., 2016).

Any LCMS tree pixel that fell over CDL non-tree crop (CONUS only) or water in respective water masks (all study areas) was recoded to non-tree. A minimum mapping unit, defined as tree clumps greater than 4 pixels (30x30 m) in size, using queen's case of adjacency (9 neighbor), was used to identify salt and pepper type tree clumps outside of urban areas in the LCMS tree mask. Any clump of tree pixels less than or equal to 4 pixels was recoded to non-tree; however, non-tree clumps less than or equal to 4 pixels were not recoded to tree.

TCC pixels that intersected non-tree pixels in the non-urban non-tree mask, were set to 0% TCC (figure 8; figure 9). For coastal Alaska, trees found in the LCMS Land Cover tall shrub class were omitted from the initial LCMS non-tree masks. The 78th tau percentile was used to separate treed from non-treed pixels in the LCMS Landcover tall shrub class in coastal Alaska.

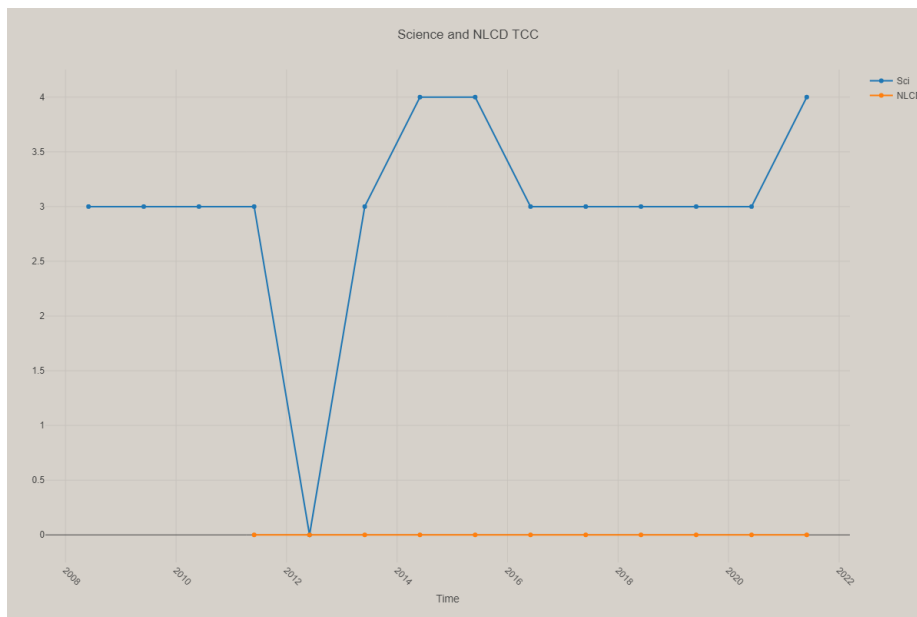


Figure 8. An example pixel where the non-urban non-tree mask filter was applied. As all years fell within the non-urban non-tree mask, the TCC values in the Science TCC product were set to 0 in the NLCD TCC product.

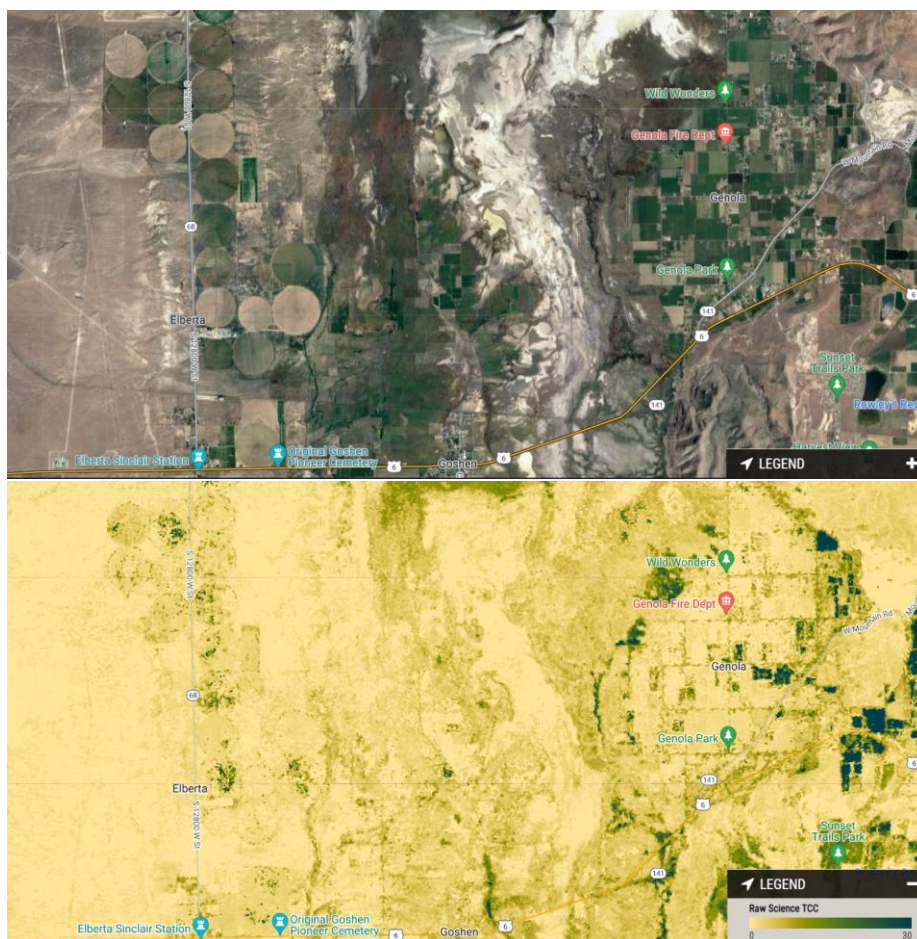




Figure 9. Example area illustrating the results of the non-urban not-tree minimum mapping unit filter. The top image shows high resolution satellite imagery for context. The middle image shows the Science TCC product. The bottom image shows the NLCD TCC product. Notice how many agriculture and wetland areas had somewhat high TCC values in the raw Science TCC product. These are then all set to 0 in this filtering step.

Urban Non-Tree TCC Masking

For urban areas, as defined above, we used a different routine to filter unplausible false positive TCC values. No minimum mapping unit over a tree mask was used as part of the urban filter routine. First, we applied the same agriculture (CONUS only) and water masks (all mapping areas) used for each year in the non-urban masking routine. Next, we used the simulated model error statistic referred to as tau, following guidance from Coulston et al. (2016; see Modeling section above) to identify pixels insignificantly different from 0% TCC. This method simulates a distribution of error values from our reference data. For the different study areas different tau percentile values were chosen. The tau percentiles chosen were the 87th percentile value for the CONUS, 70th percentile value for coastal Alaska, 90th percentile for PRUSVI, and the 90th percentile for Hawaii. This value generally balanced commission and omission of masking non-tree areas in visual exploration of different percentile values. This tau percentile value is then multiplied 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 0% TCC as follows (figure 10):

$$\left(\text{mean}(TCC_{\text{predicted}}) - (\text{standardDeviation}(TCC_{\text{predicted}}) * \text{tau}) \right) \leq 0 \rightarrow 0\%TCC$$

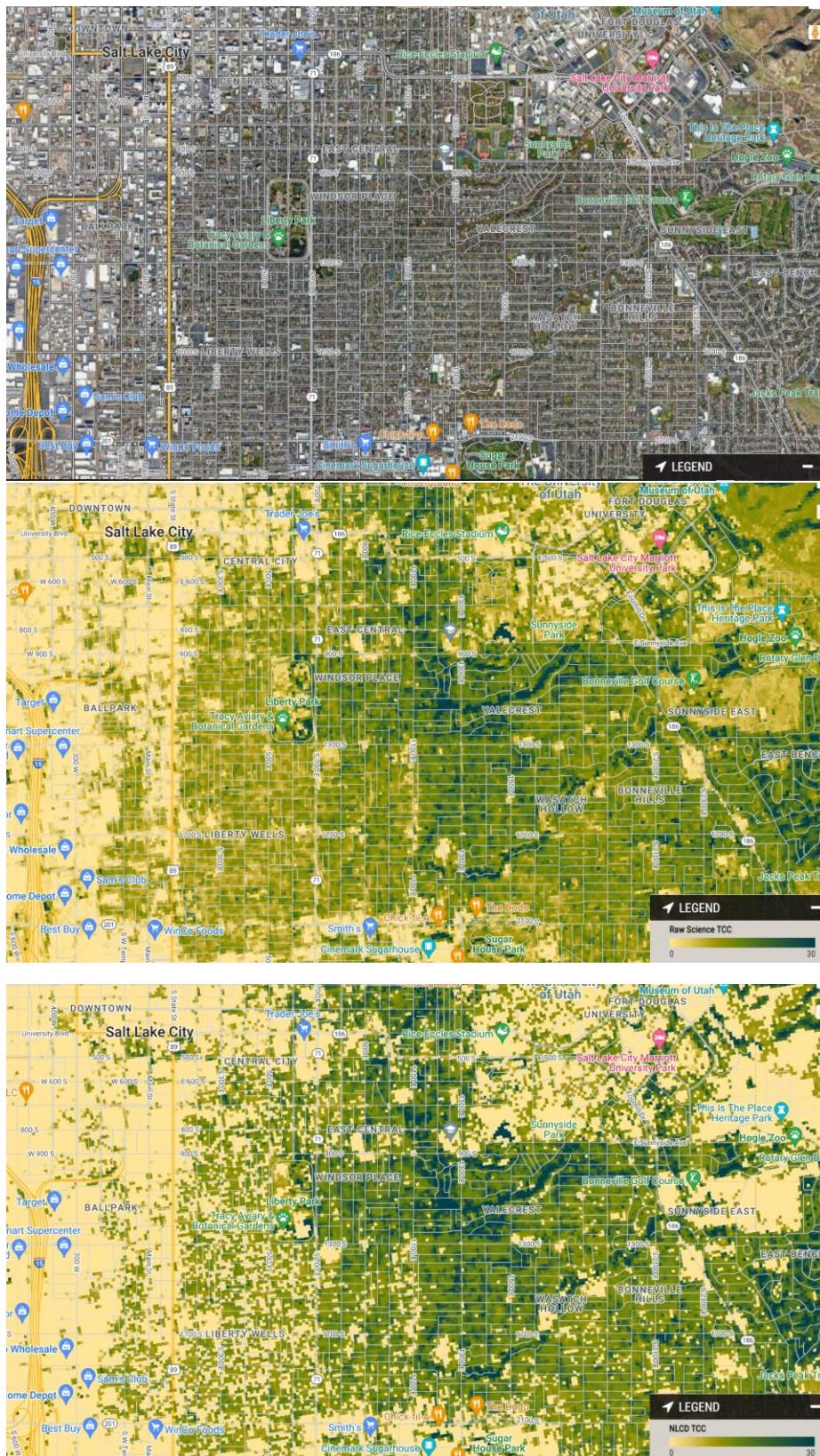


Figure 10. Urban area example from Utah, illustrating the impact of the urban low confidence filtering step. The top image shows high resolution satellite imagery for context. The middle image shows the mean TCC predictions. The bottom image shows

the filtered and masked NLCD TCC product. Notice how many higher density developed areas had somewhat high TCC values in the unfiltered TCC product. These are then all set to 0 in this filtering step.

Temporal Filtering

Last, we remove small inter-annual changes by only allowing for changes greater than $\pm 10\%$ TCC each year, starting with a 2008-2010 median value. This filtering step can have positive and negative impacts on the final outputs, as illustrated in figure 11, but simplifies the potential use of NLCD TCC outputs for change detection.

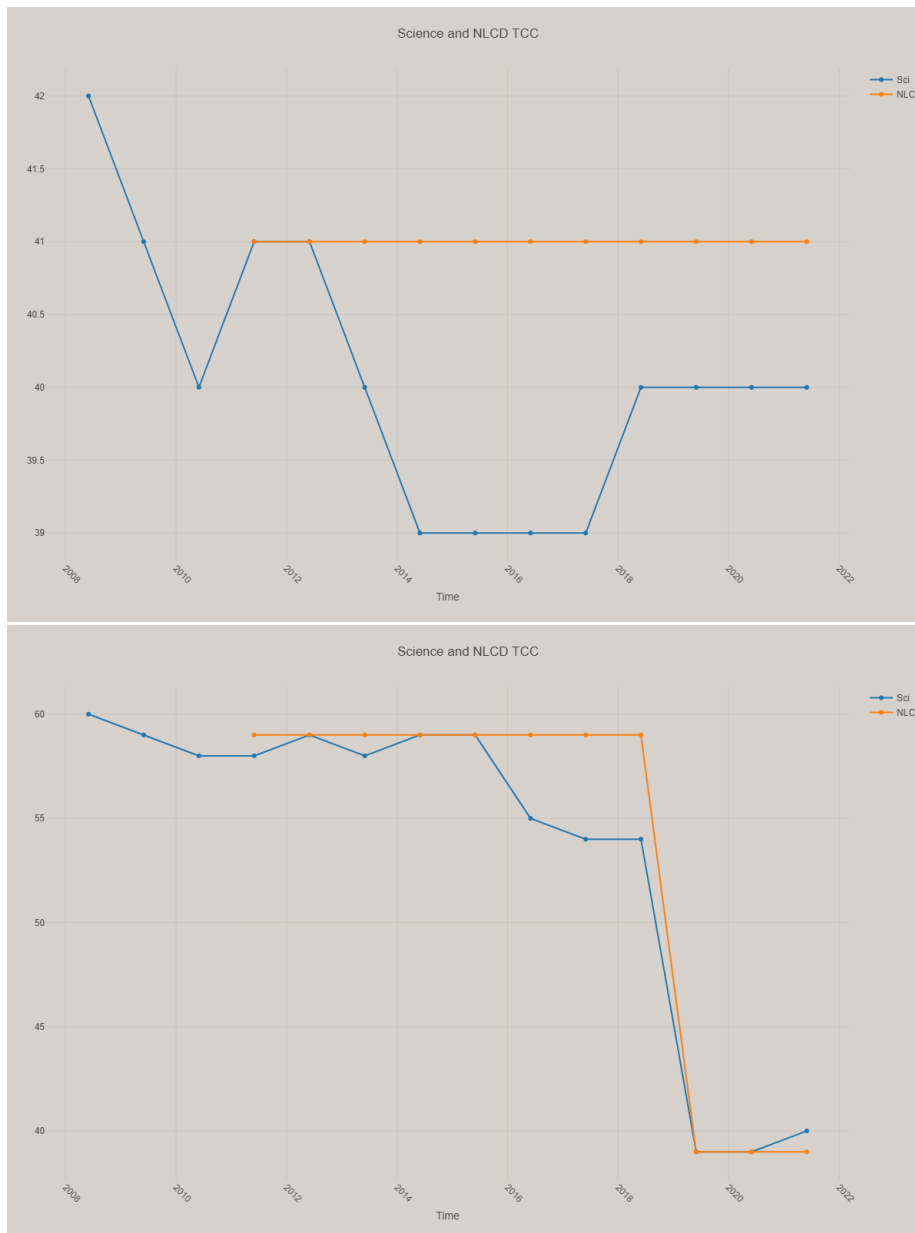


Figure 11. Comparison of the Science TCC and NLCD TCC outputs illustrating the impact of the temporal filtering routines applied. This filter is designed to prohibit any inter-annual change less than $\pm 10\%$ TCC in the final NLCD TCC products. The top example illustrates the type of small inter-annual 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).

This post-processing workflow was applied to 2011, 2013, 2016, 2019, and 2021 Science TCC outputs to produce the final NLCD TCC product suite. The NLCD TCC will now match 54 percent of the years of the USGS NLCD products, an improvement from only a 20 percent match for the 2016 release.

Map Error Assessment

We conducted an independent error assessment over the 2011 NLCD TCC v2021.4 map output. For CONUS, 30 percent of our response data was withheld from model calibration to be used for assessing error. Our 2011 CONUS FIA reference data had varying spatial densities (due to different reporting needs from FIA data, different re-measurement frequencies (Reams et al., 2005), and 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 12). 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.

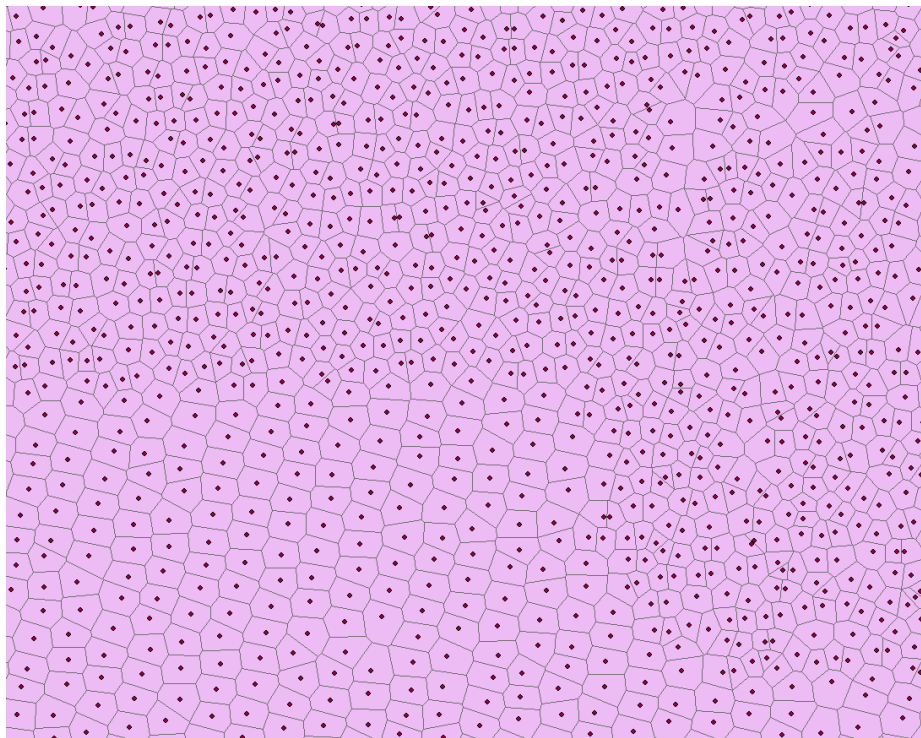


Figure 12. Fuzzed plot locations for an area with their computed Thiessen polygons shown in light gray. The area of each polygon around each plot was then used to weight the plot. This way a plot in a lower density area would receive a higher weight.

We then computed the weighted root mean squared error (RMSE) and mean absolute error (MAE) using the response TCC value as the truth, and the final NLCD TCC 2011 map value at the pixel coincident with the subplot center of the actual FIA plot location. The 2011 CONUS-wide NLCD TCC map RMSE is 13 percent TCC and the MAE is 8 percent TCC.

For coastal Alaska, 20 percent of our response data was withheld from model calibration to be used to assess error. Since all the FIA data collected for coastal Alaska occurred on the same FIA grid system, the spatial density was sufficiently consistent to negate the need for the Thiessen polygon-based weighting method used in CONUS. The RMSE was computed directly using the holdout photo-interpreted TCC calls and the predicted TCC values. The RMSE for coastal Alaska was 19 percent.

For Hawaii and PRUSVI, an independent random sample of 325 and 324 plots were used to calculate RMSE, respectively. The RMSE for PRUSVI was 21 percent and the RMSE for Hawaii was 21 percent.

Useful Resources

- [NLCD TCC Downloads](#)
- [Raw Science TCC Downloads](#)
- [ESRI Image Services](#)
- [TCC Contact Information](#)

References

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
<https://doi.org/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(2020), 18 p. <https://doi.org/10.1016/j.rse.2019.111356>
- Chastain, R., Housman, I., Goldstein, J., Finco, M., and 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(2019), 274–285.
<https://doi.org/10.1016/j.rse.2018.11.012>
- Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., and Brewer, C.K. (2012). Modeling percent tree canopy cover—A pilot study. *Photogrammetric Engineering and Remote Sensing*, 78(7), 715–727. <https://doi.org/10.14358/PERS.78.7.715>.
- Coulston, J.W., Blinn, C.E., Thomas, V.A., and Wynne, R.H. (2016). Approximating prediction uncertainty for random forest regression models. *Photogrammetric Engineering & Remote Sensing*, 82(3), 189–197. <https://doi.org/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.
- 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(2017), 379–390.
<http://doi.org/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. Morin, S. Randall, G.C. Liknes (Comps.), *Moving from status to trends: Forest Inventory and Analysis (FIA) symposium 2012 December 4-6, Baltimore MD (pp. 237-241)*. (General Technical Report NRS-P-105). U.S. Department of Agriculture, Forest Service, Northern Research Station. Newtown Square, PA.
https://www.fs.fed.us/nrs/pubs/gtr/gtr_nrs-p-105.pdf
- 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(2017), 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

Hansen, M.C., and Loveland, T.R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment*, 122(2012), 66-74.
<https://doi.org/10.1016/j.rse.2011.08.024>

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

Housman, I.W., Campbell, L.S., Heyer, J.P., Goetz, W.E., Finco, M.V., Pugh, N., and Megown, K. (2022). *US Forest Service Landscape Change Monitoring System Methods Version 2021.7* (GTAC-10252-RPT3). U.S. Department of Agriculture, Forest Service, Geospatial Technology and Applications Center, Salt Lake City, UT. 24 p. <https://doi.org/10.13140/RG.2.2.19965.23524>

Jean-Francois Pekel, Andrew Cottam, Noel Gorelick, Alan S. Belward, High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418-422 (2016). (doi:10.1038/nature20584)

Kennedy, R.E., Yang, Z., and 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.

Kennedy, R.E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W.B., and Healey, S. (2018). Implementation of the LandTrendr algorithm on google earth engine. *Remote Sensing*, 10(5), 691.

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

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

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., and 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. <https://doi.org/10.1016/j.rse.2009.12.018>

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., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(2011), 2825-2830. <https://doi.org/10.48550/arXiv.1201.0490>

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., Armston, J., Dubayah, R., Blair, J.B., Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253(2021), 112-165. <https://doi.org/10.1016/j.rse.2020.112165>

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>

Reams, G.A., Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A., and Moisen, G.G. (2005). *The forest inventory and analysis sampling frame* (General Technical Report SRS-80 pp. 21-36). U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC. https://www.srs.fs.usda.gov/pubs/gtr/gtr_srs080/gtr_srs080-reams001.pdf

Ruefenacht, B., Heyer, J., Johnson, V., Goetz, W., Bender, S., Schleeweis, K., Megown, K. (2022). *Forest Service Tree Canopy Cover Mapping: 2016 Product Suite and Methods* (GTAC-10264-RPT1). U.S. Department of Agriculture, Forest Service, Geospatial Technology and Applications Center, Salt Lake City, UT. 47 p. https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/docs/TCC_2016_MethodsReport_2022-08-16.pdf

U.S. Geological Survey. (2019). USGS 3D Elevation Program Digital Elevation Model [Online]. Available at https://developers.google.com/earth-engine/datasets/catalog/USGS_3DEP_10m

Vogeler, J.C., Braaten, J.D., Slesak, R.A., and 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(2018), 363-374. <https://doi.org/10.1016/j.rse.2018.02.046>

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., and 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. <https://doi.org/10.3390/rs6087424>

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., & Xian, G. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146(2018), 108-123. <https://doi.org/10.1016/j.isprsjprs.2018.09.006>

Zhu, Z., and Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118(2012), 83-94. <https://doi.org/10.1016/j.rse.2011.10.028>

Zupanc, A. (2017) *Improving Cloud Detection With Machine Learning*. <https://medium.com/sentinel-hub/improvingcloud-detection-with-machine-learningc09dc5d7cf13>