California Forest Observatory

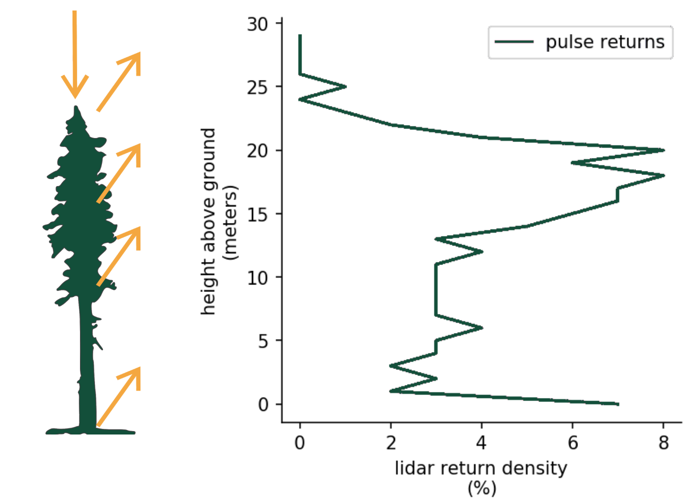
Data description

vegetation structure & fuels

April 2020 **Salo Sciences, Inc.**

The California Forest Observatory is a data-driven forest monitoring system. It maps the drivers of wildfire behavior across the state—including vegetation fuels, weather, topography & infrastructure—from space. This document describes the primary datasets produced by the Observatory: the vegetation fuels metrics. These data were derived from two data sources: airborne lidar and satellites.

A brief primer on vegetation mapping



Airborne lidar is a laser scanning technology used to map patterns of forest structure—like tree height, canopy cover, or canopy complexity. Infrared laser pulses are sent from an aircraft to the forest, and most of this energy reflects off the top of the canopy. The time between when the pulse is sent and when it returns is used to measure distance to the tree.

But not all of the energy reflects off the top of the canopy. Because of the wavelength used, some energy transmits through and reflects off the branches, leaves, and ground beneath the canopy. By measuring the timing of energy returned at these different layers, we can construct a detailed view of forest structure.

About the data

Each forest structure metric was derived directly from airborne lidar data, hosted by the USGS **3D Elevation Program**. However, these data are only available for a small fraction of California's 423,970 km² area. To overcome this, we trained deep learning models—a form of pattern recognition—to identify these forest structure patterns in satellite imagery, then mapped each metric statewide.

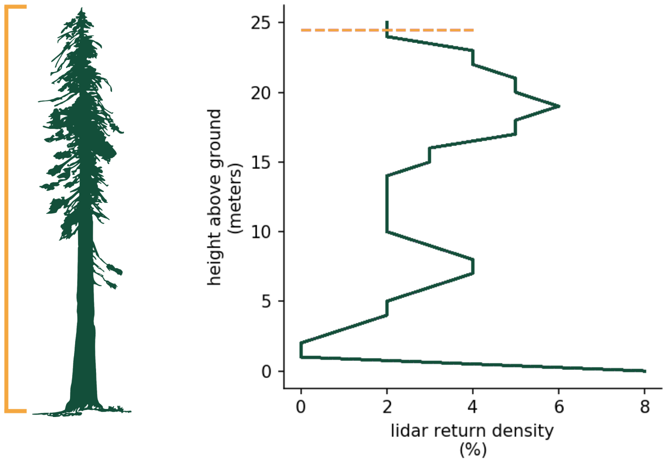
These algorithms are of the **U-net** family of neural network architectures that perform pixel-wise regression and classification tasks. The satellite data includes imagery from Sentinel-1 C-band radar sensors and Sentinel-2 multispectral sensors at 10 m spatial resolution, collected in Fall 2019. Future versions will include imagery from PlanetScope multispectral sensors at 3 m resolution.

California Forest Observatory data are stored and provided in UTM 10 North coordinate reference system (EPSG: 32610) and are available in single band GeoTIFF (**.tif**) or multi-band Landscape file (**.lcp**) raster formats.

All models are validated with at least 15,000 km^2 of airborne lidar data not used in model training. We evaluate model performance with three metrics: r^2, mean absolute error (MAE) and root mean squared error (RMSE). r^2 computes the proportion of variance in forest structure metrics (the response variables) predicted by the satellite data (the features). For canopy cover, an r^2 score of 0.91 can be interpreted to say that the model correctly explains 91% of the spatial cover patterns. MAE computes the average difference between each prediction metric and its corresponding observation. For canopy height, an MAE of 1.97 meters can be interpreted to say that tree height predictions are, on average, within 2 meters of observed values. RMSE computes the dispersion of prediction errors. It's similar to MAE, but increases non-linearly as prediction errors increase (or, RMSE penalizes large differences between predictions and observations). A large difference between RMSE and MAE indicates a high degree of dispersion—that a large number of predictions deviate significantly from observations.

Canopy fuel metrics

Canopy height



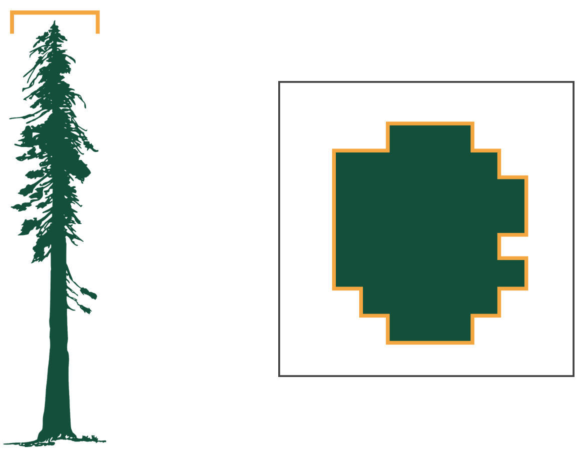
Units: meters **| Min**: 0 **| Max**: 80

Model performance

- r^2: 0.86
- **MAE**: 1.97 m
- **RMSE**: 3.61 m

Description: The distance between the ground and the top of the canopy. Canopy height is a proxy for aboveground biomass and the amount of foliage that may be consumed in a canopy fire.

Canopy cover



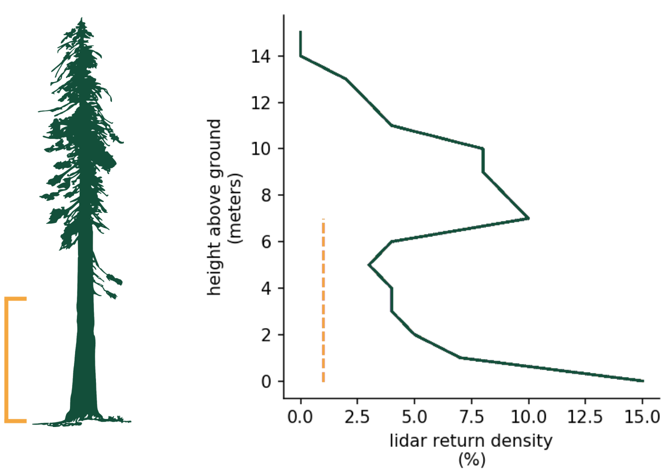
Units: percent **| Min**: 0 **| Max**: 100

Model performance

- r^2: 0.91
- **MAE**: 7.0 %
- **RMSE**: 12.2 %

Description: The horizontal cover fraction occupied by tree canopies. Maps community type & fire regime, as well as available habitat for tree-dwelling species.

Canopy base height



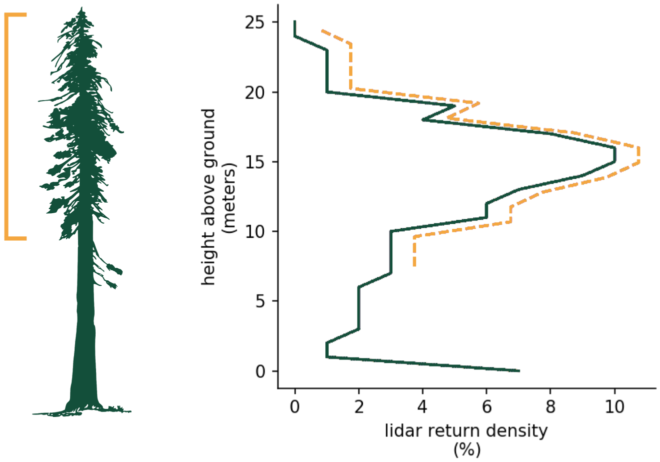
Units: meters **| Min**: 0 **| Max**: 30

Model performance

- r^2: 0.70
- **MAE**: 1.47 m
- **RMSE**: 2.53 m

Description: The distance between the ground and the lowest branches in the canopy. Predicts whether a surface fire transitions to a canopy fire.

Canopy bulk density



Units: kilograms per cubic meter | **Min**: 0 | **Max**: 0.45

Model performance: unavailable

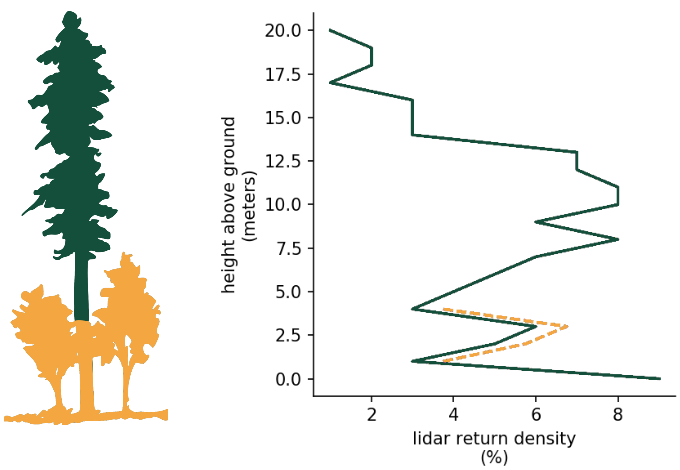
Description: The mass of available fuel that burns in a canopy fire—typically the leaves and small branches— divided by the volume of the crown. There are no performance metrics associated with CBD because it isn't directly measured. We calculate it as:

$$CBD = LMA \cdot LAI \div CV$$

Where LMA is leaf mass per area (kg m^{-2}), LAI is leaf area index (m^2 m^{-2}; see *vertical layers* below), and CV is canopy volume (m; $CV = canopy\ height - base\ height$).

Additional forest structure metrics

Ladder fuel density



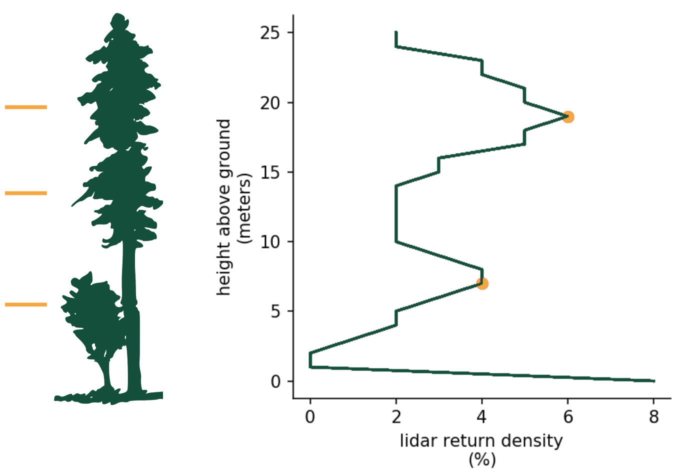
Units: percent

Model performance

- r^2: 0.78
- MAE: 2.8 %
- RMSE: 4.79 %

Description: The proportion of surface fuels in the understory. Calculated as $\#returns_{1-4m} \div \#returns_{all}$. Maps where surface fires can transition to canopy fires.

Vertical layer count



Units: count

Model performance

- r^2: 0.765
- MAE: 0.33
- RMSE: 0.49

Description: The number of distinct vertical canopy layers. Vertical layer count is a proxy for leaf area index, and maps canopy complexity.

Notes, notifications & issues

- California Forest Observatory data were developed to improve our understanding of forest structure and to improve wildfire modeling. The data are provided in two formats: as single band GeoTIFF files and as multi-band Landscape files, which were processed to match the LANDFIRE data format.
- For the GeoTIFF files, we masked all permanent water bodies as NoData. Anomalous predictions, due to clouds, snow & shadows, were also set to NoData.
- Landscape files include the LANDFIRE 2016 (Remap / LF 2.0.0) 40 Scott & Burgan Fire Behavior Fuel Models as the surface fuel model. These data were acquired from the LANDFIRE data distribution site and resampled to 10 m resolution to match the canopy fuel models. Topographic data (elevation, slope, aspect) are derived from the USGS National Elevation Dataset 1/3 arc-second (~10 m) data product.
- Since LANDFIRE doesn't support a NoData value, all NoData pixels in canopy fuel metrics were set to 0 in the Landscape files. (e.g., canopy cover was set to 0 in all NoData locations). Topographic data and surface fuel model remain unaltered.
- In the Landscape file, each canopy fuel metric was set to 0 in non-forest areas, including herbaceous/shrub systems, non-burnable types like urban, barren, snow and ice, and non-orchard agriculture.
- Model performance in agricultural and urban areas is likely worse than in forested areas. Though ag. was included in model training, the time mismatch between the lidar and satellite image acquisitions appears to have created high prediction errors, as ag. production and turnover changes so rapidly. In urban areas, some buildings were misclassified as vegetation and occasionally appear in CFO data as very tall trees. They are not.
- All satellite imagery used for model training and for prediction were collected in fall between September and November. In the 2019 data, early season snowfall at high elevations led to misclassifying snow and ice as tall vegetation. They, too, are not.
- To ensure timely data delivery, we deployed a shortcut to calculating leaf mass per area (LMA) as part of our Canopy Bulk Density calculation. We referenced mean LMA values for evergreen and deciduous species, from **de la Riva et al**, and estimated an approximate proportion of evergreen/deciduous species on a per ecoregion basis to calculate an estimate of mean per-ecoregion LMA values. LMA was then scaled by a factor of `1.556` to include the mass of small branches, per allometry from **FuelCalc**. In future versions we plan to derive LMA on a per-pixel basis directly from multispectral imagery through radiative transfer model inversion. Don't ask how, unless you're *really* curious.