Supplementary Information - Remotely Sensed Predictors of Conifer Tree Mortality During Severe Drought

P. G. Brodrick\textsuperscript{1} and G. P. Asner\textsuperscript{1}

\textsuperscript{1}Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA
Corresponding author: Phil Brodrick (brodrick@alumni.stanford.edu)
Conifer Habitats

To focus specifically on conifers, where we believe major direction shifts in NDVI (which we use to indicate mortality) are less likely to be phenological, we used the CALFIRE FVEG database to delineate conifer forest cover [CALFIRE FRAP, 2015]. We selected only pixels where the California Wildlife Habitat Relationships (CWHR) system indicated a conifer habitat (i.e. a conifer community), with no oaks listed as primary species [CDFW, 2014]. The set of conifer communities used, and the dominant species within each, are shown in Table S1.

Table S1. Conifer communities included, based on FVEG California Wildlife Habitat Relationship (CWHR) designations [CALFIRE FRAP, 2015].

| Conifer Community (CWHR Habitat) | CWHR | Dominant Species |
|-----------------------------------|------|-----------------|
| Closed-Cone Pine-Cypress          | CPC  | Tecate, Cuyamaca, Foothill Pine |
| Eastside Pine                    | DPN  | Ponderosa Pine, Jeffrey Pine, White Fir |
| Jeffrey Pine                     | JPN  | Ponderosa Pine, Coulter Pine, Sugar Pine |
| Juniper                          | JUN  | White Fir, Jeffrey Pine, Ponderosa Pine |
| Klamath Mixed Conifer            | KMC  | White Fir, Douglas Fir, Ponderosa Pine |
| Lodgepole Pine                   | LPN  | Aspen, Mountain Hemlock, Red Fir |
| Montane Hardwood-Conifer         | MHC  | Ponderosa Pine, Douglas Fir, Incense Cedar |
| Ponderosa Pine                   | PPN  | White Fir, Incense Cedar, Coulter Pine |
| Red Fir                          | RFR  | Noble Fir, White Fir, Lodgepole Pine |
| Redwood                          | RDW  | Sitka Spruce, Grand Fir, Douglas Fir |
| Sierran Mixed Conifer            | SM C | White Fir, Douglas Fir, Ponderosa Pine |
| Subalpine Conifer                | SCN  | Englemann Spruce, Subalpine Fir, Mountain Hemlock |
| Undetermined Conifer             | CON  | -               |

Deep Learning Model for CWC Time Series

The methodology used in this work was original developed in Asner et al. [2016], but has been altered slightly in this work to account for an additional campaign of data collection. Here we summarize the Asner et al. [2016] methodology and highlight the differences. As briefly described in the main text, we modeled a CWC time series for the July-August time periods of 2010-2016 using a deep learning model. This deep learning model is an example of a supervised machine learning algorithm, which is a statistical method of linking a series of input features (in this case an environmental dataset) to a response variable (in this case CWC). A deep learning model was selected for this work after comparing the results of generalized linear models, random forest regressions, and gradient boosting regressions. Deep learning models are based on a series of linked ‘neurons’, which together with a series of nonlinear functions determine how features interact with one another. Multi-layer deep learning models allow for the encapsulation of sophisticated nonlinear relationships between features and the response variable.

Key to the performance of the deep learning model is the set of input features that the model is trained on to correlate with CWC. We use a combination of satellite and terrain data, specifically elevation, slope, aspect, and relative elevation data derived from the USGS National Elevation Dataset (available at https://nationalmap.gov/), distance to road and road density derived from the U.S. Census Bureau of Topologically Integrated Geographic Encoding and
Referencing, 2014 (available at ftp://ftp2.census.gov/geo/tiger/), potential solar insolation derived from elevation data using SAGA GIS Potential Insolation module (SAGA-Python v0.37), and finally Landsat-8 data processed to surface reflectance by the USGS, and mosaicked over the July-August time period to remove cloud cover.

We train the deep learning model on all locations with airborne CWC data collected in both 2015 and 2016. The features used in the training data are as described in the previous paragraph, with the surface reflectance data used matching each year. In total, we have 29.75 million data points with which to train the model, of which 10% from each year are withheld to assess the model performance. The deep learning model consists of 4 internal layers of 200 neurons each, with rectifier linear unit (ReLU) activation functions, and was built using a Theano framework [The Theano Development Team et al., 2016]. The r-squared values for the comparison between the 2015 and 2016 predicted CWC and the holdout sets were .89 and .88 respectively, and the root mean squared errors were 0.34 and 0.35 respectively.

After the model is trained, we then substitute the surface reflectance data from 2015 and 2016 with surface reflectance data from each year between 2010 and 2016 (inclusive). Surface reflectance data from 2012 is skipped because only Landsat-7 data are available during that year, and significant artifact is present in that dataset. In Asner et al. [2016], we also applied the model to areas outside of the flight line, and consequently discussed several additional masking features – however, in this work this was not necessary as our sole concern was areas with flight data collected in both 2015 and 2016.

**Absolute CWC loss**

Figures S1 and S2 provide additional correlations between the absolute loss in CWC between 2014 and 2015 and the subsequent increase in mortality between 2015 and 2016. We chose to use percent difference rather than absolute difference in the main analysis, but Figure S1 shows an overall similar pattern with absolute CWC loss. Some community-specific patterns in Figure S2 do show different relationships, but not in a consistent enough manner to make absolute loss in CWC a preferable indicator of mortality to percent difference.
**Figure S1.** Relationship between the absolute loss in canopy water content between 2014 and 2015 and the fractional mortality increase between 2015 and 2016. The dot size in the top plot shows the relative quantity of data within each bin (approximately 296,000 data points were used in total). Controlled sampling, shown by the red dots, was performed using Latin Hypercube Sampling over space and conifer community. The conifer community representation in the bottom plot shows the fraction of each community within the corresponding pixel in the top plot.

**Figure S2.** Relationship between the absolute loss in canopy water content between 2014 and 2015 and the fractional mortality increase between 2015 and 2016 for each of the five most common conifer communities in the sampled data. Dot size shows the relative quantity of data at each point (approximately 218,000 data points were used in total).
References

Asner, G. P., P. G. Brodrick, C. B. Anderson, N. Vaughn, D. E. Knapp, and R. E. Martin (2016), Progressive forest canopy water loss during the 2012-2015 California drought., *Proc. Natl. Acad. Sci. U. S. A.*, 113(2), E249-55, doi:10.1073/pnas.1523397113.

California Department of Fish and Wildlife (CDFW) (2014), *Users manual for version 9.0 of the California Wildlife Habitat Relationships System and Bioview*, Sacramento, California.

California Department of Forestry and Fire Protection (2015), Vegetation (fveg) - CAL FIRE FRAP [ds1327],

The Theano Development Team, T. T. D. et al. (2016), Theano: A Python framework for fast computation of mathematical expressions,