Quantifying Drought-Induced Tree Mortality in the Open Canopy Woodlands of Central Texas

Amanda M. Schwantes^a, Jennifer J. Swenson^a, Robert B. Jackson^{a,b}

^aNicholas School of the Environment, Duke University, Box 90328, Durham, NC 27708, USA.

^bDepartment of Earth System Science, Woods Institute for the Environment, and Precourt Institute for Energy, Stanford University, Y2E2 Building, 379B, Stanford, CA 94305, USA.

Email Addresses:

Amanda M. Schwantes (amanda.schwantes@duke.edu), Jennifer J. Swenson (jswenson@duke.edu), Robert B. Jackson (rob.jackson@stanford.edu)

Corresponding author: Amanda M. Schwantes. Email: amanda.schwantes@duke.edu.

Telephone: 571-344-2883

ABSTRACT

During 2011, Texas experienced a severe drought, which caused substantial tree mortality. Drought-induced tree mortality can have significant ecological impacts and is expected to increase in many locations with climate change. This disturbance is unique because it often is limited to only subtle and diffuse changes in forest cover. Thus we developed new methods to quantify drought-driven canopy loss using remotely sensed imagery, within a Landsat scene in central Texas ($>30,000 \text{ km}^2$). First, fine-scale canopy loss maps were created by classifying 17 1-m orthophotos (each $\sim 50 \text{ km}^2$) from the U.S. National Agriculture Imagery Program. These classifications were highly correlated ($R^2=0.90$) with field estimates of canopy cover loss measured in 21 plots at 4 sites across central Texas. These fine-scale canopy loss maps were then used to calibrate and validate coarser-scale Landsat imagery. In scaling up to create regional canopy loss maps, we assembled a Landsat time-series and separated mortality pixels experiencing persistent canopy loss from pixels with only background noise by applying the Landtrendr algorithm. We then estimated percent tree canopy loss within each of these mortality pixels by comparing two models capable of handling zero-inflated continuous proportions: random forest and a zero-or-one inflated beta (ZOIB) regression model. We found that the ZOIB regression model had the highest accuracy in predicting canopy loss (mean absolute error = 5.16%, root mean square error = 8.01%). The 2011 drought caused a decrease in canopy cover within the study area, equivalent to $1,124 \text{ km}^2$ of canopy loss, $\sim 10\%$ of the $10,850 \text{ km}^2$ area of live canopy present before the drought. Our methods address the need to detect drought-induced tree mortality as extreme droughts are predicted to increase with climate change. More detailed canopy loss maps could then be used (1) to quantify potential impacts to

carbon cycling, biophysics, and species compositions and (2) to understand the factors controlling tree mortality, now and in the future.

Keywords: Drought; Tree mortality; Canopy loss; Disturbance; Die-back; Change detection; Zero-or-one inflated beta regression models; Random forest; Landsat time-series; Landtrendr.

1. Introduction

Tree mortality due to drought is increasing worldwide (Allen et al., 2010), ranging from small but consistent increases in background mortality rates (van Mantgem et al., 2009) to extensive regional die-offs (Breshears et al., 2005). This increased mortality will likely have large ecological consequences by altering plant community composition, species distributions (Engelbrecht et al., 2007; Mueller et al., 2005; Willson, Manos, & Jackson, 2008) and carbon (Michaelian, Hogg, Hall, & Arsenault, 2011) and water cycling (Guardiola-Claramonte et al., 2011). Establishing protocols capable of detecting drought-induced tree mortality is therefore essential.

From October 2010 to September 2011, Texas experienced the most severe one-year drought since record-keeping began in 1895 (Hoerling et al., 2013). A recent modeling study suggests that the likelihood of abnormally high summer temperatures has increased in Texas (Rupp et al., 2015). However, it is difficult to quantify changes in the probability of an extreme event, and thus challenging to definitively link the 2011 drought with directional climate change. Nevertheless, global climate models predict that subtropical dry zones, which include our study area, will become drier as climate change advances (Seager, Naik, & Vecchi, 2010). Additionally, even without decreasing precipitation, warmer temperatures will increase the

frequency of droughts, affecting ecosystems faster and more intensely than in recent decades and centuries (Trenberth et al., 2014), likely leading to increased vulnerability of trees to drought stress (Allen, Breshears, & McDowell, 2015).

Currently DGVMs (dynamic global vegetation models) are being used to predict how ecosystems will respond to rising CO₂ concentrations. These models include multiple atmosphere-biosphere interactions and feedback effects. However, the mortality algorithms in DGVMs are not typically tested or calibrated with empirical data due to a lack of systematic observations of tree mortality at regional/global scales (McDowell et al., 2011). Therefore, defining methods capable of quantifying tree mortality regionally using remotely sensed data is essential (McDowell et al., 2015), in part because improving the predictive capability of DGVMs could aid in our ability to forecast tree mortality in the face of climate change.

Already, Landsat data have played a pivotal role in our understanding of forest disturbance, being used to observe the Earth surface for 40+ years and having been designed to achieve a balance between spatial and temporal resolutions ideal for detecting forest change at scales relevant to ecology (Cohen & Goward, 2004). Previous studies have used Landsat time-series to quantify tree mortality from disturbance events due to fire (Collins et al., 2009), insects in combination with drought (Clifford, Royer, Cobb, Breshears, & Ford, 2013; Meddens & Hicke, 2014; Meigs, Kennedy, & Cohen, 2011) and clear-cut logging (Cohen, Fiorella, Gray, Helmer, & Anderson, 1998). However, fewer studies have examined ecosystems where the disturbance's proximate cause is directly from drought.

Drought-induced tree mortality is a unique forest disturbance. It can occur broadly and simultaneously across a landscape, without geographic proximity, and unlike other disturbances such as fire and pests, there is typically no centralized starting location. Detection methods must

be capable of quantifying scattered and subtle changes in canopy cover across large regions. Stand-replacing disturbances, such as fire, are easier to detect compared to disturbances that are diffuse and heterogeneous, such as drought (McDowell et al., 2015).

Most forest change detection studies classify an entire pixel as either disturbed or undisturbed (a binary per-pixel classification); however, for drought-induced tree canopy loss most pixels are mixtures of both live and dead canopy. To add to the complexity, our study site is dominated by open canopy woodlands, with an average canopy cover of only ~46% before the drought, as measured in this study. Binary per-pixel classification approaches are often inadequate in detecting subtle changes in land cover, especially in these mixed-cover systems, for instance, where a pixel could have started with only 50% canopy cover and lost 10% during the drought.

Typically, sub-pixel estimates of canopy cover use Spectral Mixture Analysis (SMA). However, our study region is composed of diverse natural systems, and as such the number of endmembers required for SMA would be larger than the spectral resolution of Landsat. Fewer studies have explored alternatives to SMA that are still capable of detecting percent canopy loss within a pixel. Of these studies, continuous tree canopy loss is usually modeled using linear regressions (Meddens & Hicke, 2014). However, for the modeling of continuous proportions, such as canopy cover loss within a pixel, linear regressions are typically not suitable (Ospina & Ferrari, 2012). Proportions are inherently bounded on the interval from 0 to 1. Therefore, this type of data defies two assumptions of a normal linear regression: (1) the relationship is non-linear and (2) the variance is heteroskedastic, because as the mean approaches the bound, the variance approaches zero (Kieschnick & McCullough, 2003). Linear regressions are especially not suitable if a dataset is zero-inflated. Zero-inflation occurs when a dataset includes an

excessive number of zeros, and is common in most remotely sensed mortality datasets, because much of the landscape remains undisturbed. Instead, we propose two approaches that can more adequately handle zero-inflated continuous proportions: (1) a zero-inflated modeling approach using random forest, RF, (Breiman, 2001), as used in Savage et al. (2015), and (2) a zero-or-one inflated beta regression model, ZOIB (Ospina & Ferrari, 2010).

We quantify drought-induced canopy loss by first filtering mortality (canopy loss) pixels from no-change pixels (pixels with no canopy loss) based on a Landsat time-series analysis (2000-2014). Then, using Landsat data from before and after the drought, as well as auxiliary variables, we model continuous canopy loss over a region. We build and validate models using training and testing data from fine-scale orthophoto classifications that span a subset of the study area, and then use these models to predict canopy loss across the entire region. Our goal was to design a method capable of detecting drought-induced canopy loss in open canopy forests, by (1) filtering canopy loss from non-change pixels beyond simple thresholds, (2) detecting continuous canopy loss within a pixel, and (3) addressing zero-inflation.

2. Study area

The study area was located in the Edwards Plateau and Llano Uplift regions of central Texas (Fig. 1). The average annual precipitation (1980-2010) was 73 cm (Thornton et al., 2014; Thornton, Running, & White, 1997). Average annual daily mean and max temperatures (1980-2010) were 18.5 and 38.5°C, respectively. In 2011, the region received only 34 cm of rain, coupled with higher mean and maximum temperatures of 19.6 and 41.2 °C, respectively (Thornton et al., 2014).

Dominant tree species in the Edwards Plateau & Llano Uplift region include *Juniperus ashei* (ashe juniper), *Quercus fusiformis* (plateau live oak), *Quercus stellata* (post oak), *Ulmus crassifolia* (cedar elm), and *Prosopis glandulosa* (honey mesquite) (Elliott et al., 2014). Our analysis only included areas that were forested; a binary forest cover mask was created by including all ecological systems from the Texas Ecological Systems Map (Elliott et al., 2014) dominated by tree species. Growth form (shrub vs. tree) was defined using the United States Department of Agriculture (USDA) plants database (USDA Natural Resources Conservation Service, 2015). We also excluded areas influenced by fire using data from the Monitoring Trends in Burn Severity project (USDA Forest Service/U.S. Geological Survey, 2014).

Fig. 1. Study area (Landsat scene path 28/row 39) in the Edwards Plateau/ Llano Uplift region of central Texas.

3. Methods

To obtain estimates of tree canopy loss across an area $> 30,000 \text{ km}^2$ in central Texas, we first created a subsample of fine-scale canopy loss maps based on orthophotos (1-m) from the U.S. National Agriculture Imagery Program (NAIP). A Landsat time-series analysis (2000-2014) was used to filter areas with canopy loss from regions experiencing no change, using the Landtrendr algorithm. Landtrendr is a time-series change detection algorithm, specifically designed for Landsat data (Kennedy, Yang, & Cohen, 2010). Pixels containing burned areas, non-forest, or no canopy loss were excluded from the analysis. We then built models, relating the fine-scale canopy loss maps to coarse-scale Landsat imagery. These models were used to predict post-drought proportion canopy loss and live canopy across the study area at a 60-m spatial resolution (Fig. 2).

Fig. 2. Canopy loss examined across scales: (a) orthophotos (1-m) are used to create (b) fine-scale canopy loss maps (1-m), which are then validated with (c) ground observations of canopy cover loss. Then models are built relating these fine-scale canopy loss maps to (d) coarser scale Landsat and MODIS imagery. Pixels of (e) non-forest, no-change, or burned area were removed from the study area. Finally these models were used to predict (f) regional coarse-scale continuous live canopy and canopy loss, across the study area, aggregated to a spatial resolution of 60-m.

3.1. *Creating fine-scale canopy loss maps*

3.1.1. *Classification of orthophotos*

NAIP digital orthophotos (1-m) were used to create high resolution canopy loss maps (see below) to act as training data for coarser scale images. NAIP images (US Department of Agriculture, 2014) for our area contain four bands (red-R, green-G, blue-B, and near infrared-NIR) and are taken during the growing season. These images were already orthorectified to within +/- 6-m to true ground, and tiled using 3.75' x 3.75' quarter quads. Thirteen quarter quads were randomly chosen as training/testing data for this study. Only quarter quads with cloud-free images were included. To match images with field data for validation, four additional quarter quads were selected over areas with access in the field, which is restricted across the study area. Two of these four quarter quads were used exclusively for field validation, because they fell outside of the Landsat scene, but still within the natural systems that were the focus of this study: the Edwards plateau and the Llano uplift (Fig. 1).

The 17 quarter quad NAIP image pairs, one from pre-drought (2010) and one from post-drought (2012), were co-registered (each with 30 manually chosen tie points, 1st or 2nd order polynomial warping transformation, Root Mean Square Error, RMSE <1.0 m, and nearest neighbor resampling). Then normalized difference vegetation index (NDVI) and the original R,G,B,NIR bands, both from before and after the drought, were included in one image stack for each of the 17 quarter quads. A supervised classification was performed on each multi-date stack (10 bands: 2010 - R, G, B, NIR, NDVI and 2012 - R, G, B, NIR, NDVI) using the maximum likelihood algorithm. Training data were identified by manually delineating dominant features within each scene as well as features that changed through time from 2010 to 2012. Post-processing of these classifications included grouping these sub-classes into three main classes:

(1) live canopy (e.g., live trees, shrubs, barely alive/thinned canopies); (2) canopy loss (e.g., originally live canopy in 2010 and then dead canopy in 2012) and (3) non-tree cover (e.g., agriculture, shadows, water, urban, pasture, bare soil). Isolated groups of three pixels were removed and re-classified using a 3 by 3 majority filter.

3.1.2. Accuracy assessment of fine-scale canopy loss maps with field surveys

Canopy cover loss was measured in the field in four sites across the Edwards Plateau and Llano Uplift region in July 2013, two years after the drought. At each of the 4 sites, 3-6 plots were sampled for a total of 21 plots. Within each site, plot locations were selected using a stratified random sample. For each of the four fine-scale canopy loss maps, three equal area strata were assigned based on tree canopy cover loss (high, medium, & low). Thresholds unique to each image were chosen so that an equal area of the image was assigned to each of the three strata. An equal number of plots were then randomly selected within each stratum. This ensured that the entire range of canopy cover loss was sampled within a site, but that all areas had an equal probability of being selected. If a plot location contained evidence of a disturbance other than drought (i.e. fire) that could have occurred after the drought (2011) but prior to our field campaign (2013), then a new plot location within the same stratum was selected.

For each 50 x 50-m field plot, we used the line-intercept method to measure canopy cover (to the nearest 0.1 m) along four 25-m transects running from the randomized plot center in each of the four cardinal directions (N, E, S, W). Along each transect the health status of the tree or shrub was also recorded as either dead or alive. We assumed that the canopy death observed along each transect happened during the 2011 drought interval. For example, background mortality rates in east Texas are typically low ~0.7% (Moore et al., in press); therefore, very few

dead standing trees were present in these plots prior to the drought. Trees, which appeared dead for greater than 3-4 years (e.g. missing bark and medium to large branches), were not counted as dead as a result of the 2011 drought. A perpendicular metal sampling rod was used to estimate the intersection of a tree's canopy with the measuring tape. When trees were taller than the 2.5-m sampling rod, a GRS densitometer (sighting device) was used to estimate canopy cover. Only woody shrubs and trees over 1-m tall were counted. Transect length was corrected for slope, in steep terrain (slope $>5^\circ$). GPS points were collected at the plot center and later differentially corrected. The proportion of live and dead canopy along each of the four line-transects was averaged, to obtain estimates of the average live canopy and canopy loss for each plot. These ground plot measurements were then compared to the 1-m orthophoto classification corresponding to each site, where the proportions of post-drought live canopy pixels and canopy loss pixels were estimated within a 25-m radius circular buffer from each plot's center. Four of the seventeen classifications were validated with field data to test the accuracy of the fine-scale (1-m) canopy loss classifications. However, only 15 out of the 17 classifications were used in scaling up to create coarse-scale regional canopy loss maps.

3.2. Creating coarse-scale canopy loss maps: pre-processing steps

We acquired Landsat imagery from before and after the drought. Areas of no canopy loss were masked using a Landsat time-series analysis. Then, grids aligned with the 30-m Landsat pixels were placed over the 1-m fine-scale tree canopy loss maps. Proportions of live, canopy loss, and non-tree cover were then extracted for each 30-m grid cell. These percent canopy loss observations were then used to calibrate the Landsat imagery overlying the footprints of the orthophoto classifications.

3.2.1. Pre-processing of Landsat imagery

Pre-drought and post-drought images were selected for dates on July 10, 2009 & July 5, 2013, respectively. We acquired atmospherically corrected images from the Landsat surface reflectance product (United States Geological Survey, 2015). Through visual inspection, 1-m fine-scale canopy loss maps, which showed imperfect registration with the Landsat scenes, were manually co-registered to the Landsat images to achieve a RMSE < 0.5 of a Landsat pixel.

3.2.2. Filtering canopy loss from no-change pixels: Landtrendr algorithm

Some parts of our study region experienced no canopy loss; therefore, the dataset was highly zero-inflated, which can limit model performance. To concentrate the analysis only on pixels that actually contained canopy loss, a Landsat time-series analysis was conducted to distinguish pixels with change in canopy cover from pixels with noise resulting from year to year atmospheric effects, geometric mis-registration, and/or shadowing or other illumination differences (Kennedy et al., 2010). Change vs. unchanged areas have also been delineated in other studies by setting thresholds, using the difference in a vegetation index between two time-periods (Rogan, Franklin, & Roberts, 2002). We chose to use a more rigorous time-series analysis to identify areas undergoing a persistent change in canopy cover.

We selected late-summer Landsat 5, 7 & 8 images (path 28/row 39) from 2000 to 2014. Images with clouds or gaps attributable to scanline errors (Landsat 7) were filled in by other imagery to obtain cloud-free aggregates from 2000 to 2014. By late summer at our study site, herbaceous vegetation has typically senesced, thereby allowing for better differentiation between tree mortality and grass/cropland senescence. We calculated NDVI for each cloud-free image

composite of late-summer Landsat imagery from 2000-2014. We then applied the Landtrendr temporal segmentation algorithm to this NDVI time-series. The segmentation fitting protocol is capable of capturing both long-term trends and abrupt events (Kennedy et al., 2010). Following segmentation, we extracted maps of segments containing the most recent disturbance in the time-series. If the year 2011 or after occurred within that segment, then the pixel was classified as having experienced canopy loss. No-change pixels were considered to have 0% canopy loss, and canopy loss pixels were used in the next steps to create continuous canopy loss maps. Isolated pixels of no-change that were surrounded by canopy loss pixels were removed and reclassified as canopy loss using a 3 by 3 majority filter, applied twice. Across the study area, 44% of the pixels were considered no change.

Typically, Landtrendr is parameterized to achieve a balance of false positives (no-change pixels that were considered change), and false negatives (change pixels that were considered no-change); however, we evaluated several different parameter values and selected the final parameterization that strongly minimized false negatives at the expense of also selecting for more false positives. We did this to ensure that pixels containing canopy loss were not removed as no-change, because drought-induced tree mortality can cause subtle changes in canopy cover. Furthermore, both the zero inflated modeling approach using RF (Savage et al., 2015) and ZOIB regression models are capable of handling some zero-inflation, and thus these false positives were filtered out in subsequent steps.

3.2.3. Scaling up to create regional continuous canopy loss maps

Proportional canopy loss, post-drought live canopy, and non-tree cover from the 15 fine scale canopy loss maps were extracted for each 30-m Landsat pixel. The data were then

aggregated to 60-m to minimize the effect of misalignment errors between the orthophotos and Landsat. Data aggregation occurred by taking the mean value of all 30-m pixels within the 60-m grid cell. We decided on an aggregation level of 60-m by testing multiple aggregation levels (Fig. S1, supplementary materials). The 60-m spatial resolution decreased the error rate significantly, while only marginally increasing the spatial resolution of Landsat. Only forested and non-burned pixels with greater than zero canopy loss were considered in the analysis. The proportions of post-drought canopy loss and live canopy within each 60-m grid (75,716 cells total) were then used to calibrate and validate the coarse-scale 60-m aggregated imagery, within the footprint of the orthophoto classifications.

3.3. Creating coarse-scale canopy loss maps: variable selection and modeling

To predict canopy dynamics across a broad area based on the fine-scale maps, we tested the predictive power of two models (RF and ZOIB) by comparing two designs for the training and testing data: (1) where testing and training datasets were close in proximity and (2) where the testing data was spatially independent of the training data. Many studies use either plot data or fine-scale imagery in scaling up to regional estimates; however, often this training/testing data is from only a spatial subset, and yet used to validate the predictive power of the model across an entire region. We explored using spatially independent testing data, which proved to be a more accurate representation of the predictive capability of a model (González-Roglich, 2016). We built models relating the canopy loss observed in the orthophoto classifications to the Landsat data and auxiliary explanatory variables within the footprints of the orthophotos, which cover approximately 2.4% of our study area. We then used the final models to create post-drought continuous (1) canopy loss and (2) live canopy maps across the entire $>30,000 \text{ km}^2$ study area.

3.3.1. Identifying explanatory variables: vegetation indices and auxiliary variables

Ten vegetation and water indices, including tassell cap transformations, were calculated for the pre-drought (2009) and post-drought (2013) Landsat scenes as well as their difference over time (2013-2009) (Table 1). The tassell cap transformation compresses the six original Landsat bands to create three orthogonal indices, brightness, greenness, and wetness (Baig, Zhang, Shuai, & Tong, 2014; Crist, 1985).

Table 1
Summary of vegetation indices considered in the analysis

Index	Formula	Citation
Normalized Difference Vegetation Index (NDVI)	L5: $(B4-B3)/(B4+B3)$ L8: $(B5-B4)/(B5+B4)$	(Tucker, 1979)
Normalized Difference Water index (NDWI)	L5: $(B4-B5)/(B4+B5)$ L8: $(B5-B6)/(B5+B6)$	(Wilson & Sader, 2002)
Normalized Burn Ratio (NBR)	L5: $(B4-B7)/(B4+B7)$ L8: $(B5-B7)/(B5+B7)$	(van Wagtendonk, Root, & Key, 2004)
Greenness Normalized Difference Vegetation Index (GNDVI)	L5: $(B4-B2)/(B4+B2)$ L8: $(B5-B3)/(B5+B3)$	(Gitelson, Kaufman, & Merzlyak, 1996)
Ratio SWIR/NIR (B5/B4)	L5: $B5/B4$ L8: $B6/B5$	(Vogelmann & Rock, 1988)
Red Green Index (RGI)	L5: $B3/B2$ L8: $B4/B3$	(Coops, Johnson, Wulder, & White, 2006)
Difference Vegetation Index (DVI)	L5: $(B4-B3)$ L8: $(B5-B4)$	(Tucker, 1979)
Tasseled cap brightness (TCB), greenness (TCG), wetness(TCW)	See citations	(Baig et al., 2014; Crist, 1985)

In addition to the Landsat vegetation indices, six other sets of explanatory variables were considered: (1) percent forest cover, (2) location: latitude & longitude in UTM coordinates, (3)

land surface temperature from MODIS, (4) elevation, (5) slope & aspect, and (6) magnitude of change from the time-series analysis using Landtrendr. Percent forest cover was obtained from National Land Cover Database (Homer et al., 2015). To calculate land surface temperature (LST), we used the 8-day 1-km daytime LST dataset from the MODIS/Aqua satellite (MYD11A2; collection 5 (Wan, 2008)). Pre-processing included the following: (1) removal of poor quality data using the QC flags (data with LST error rates greater than 2 K or emissivity errors greater than 0.02 were excluded), (2) aggregation of the 8-day product to warm season averages (May-September), and (3) resampling to the 60-m aggregated Landsat grid. We used the 2011 LST difference from the historical average (calculated from 2002-2010) in our models.

Elevation, slope and aspect were all computed using the National Elevation Dataset, downloaded from the U.S. Geological Survey. Slope (ϕ) and aspect (θ) were transformed using the following equations, to account for joint effects (e.g. the influence of aspect would be stronger on steeper slopes). The first two equations include both aspect and slope, and take into account the north-south ($u1$) and east-west gradients ($u2$) of aspect. The last equation ($u3$) tests for the influence of slope independent of aspect (Clark, 1990).

$$\begin{aligned}u1 &= \cos(\theta) \sin(\phi) \\u2 &= \sin(\theta) \sin(\phi) \\u3 &= \cos(\phi)\end{aligned}$$

Each segment in the time-series analysis has an associated magnitude of change, which is computed as the difference in NDVI prior to the disturbance and NDVI after. We also used this magnitude of change estimate as an additional explanatory variable in our models. To verify that prediction outside of the training dataset was appropriate, we compared the range of values for each explanatory variable in the training dataset to the proportion of pixels across the whole study area that fell within that range (Table 2). For each explanatory variable, less than 0.03% of

337 the pixels in the study area fell outside the range used in training the model, except for the
338 covariates, elevation, latitude, and longitude, which were not used for prediction in the final
339 models.

340

Table 2

Covariates considered in modeling: ranges of values in the training dataset compared to the percent of pixels across the study area that fell outside of these ranges.

Variable	Model (A/CL) ¹	Range (training data)	% pixels in study area outside of the range ²
Vegetation indices 2009			
NDVI	CL	-0.00165 to 0.821	***
NDWI	CL	-0.506 to 0.341	**
NBR		-0.265 to 0.636	**
GNDVI		-0.0331 to 0.779	***
(B5/B4)		0.493 to 3.05	**
RGI	CL	0.766 to 1.45	**
DVI	CL	12.0 to 3642	***
TCB ³	CL	-5517 to 6585	***
TCG ³	CL	-1201 to 1715	**
TCW ³	CL	-2104 to 6917	***
Vegetation indices 2013			
NDVI	CL, A	-0.194 to 0.863	**
NDWI	CL	-0.299 to 0.405	**
NBR	A	-0.264 to 0.689	***
GNDVI		-0.223 to 0.800	***
(B5/B4)		0.424 to 1.86	**
RGI	CL, A	0.632 to 1.58	**
DVI	CL, A	-253 to 3939	***
TCB	CL, A	778 to 14701	***
TCG	CL, A	-467 to 2724	***
TCW	CL, A	-5240 to 298	***
Auxiliary variables			
% tree cover	A	0.250 to 96.0	***
X		360900 to 566070	1.0%
Y		3263100 to 3416220	11%
MODIS LST	CL	2.93 to 11.3	***
Difference			
Elevation		285 to 721	2.0%
U1		-0.545 to 0.674	***
U2		-0.510 to 0.541	**
U3		0.697 to 1.00	***
Time series	CL, A	0 to 626	***

¹Indicates if the explanatory variable was used to model Canopy Loss (CL) or Alive (A) cover

²Percent of pixels predicted within the study area that fell outside of the range of the explanatory variables used to train the model

³Indicates that the range of the difference (2013-2009) is provided instead of 2009 values

*** < 0.01 %   ** < 0.03 %

3.3.2. Modeling approaches: RF vs ZOIB

Two modeling approaches were used to detect canopy loss within each 60-m grid cell: (1) a zero-inflated modeling approach using RF and (2) a ZOIB regression model. For both modeling approaches, the percent canopy loss in a pixel area was modeled separately from percent post-drought live canopy, with percent non-tree cover computed as the remaining cover not already accounted for by either canopy loss or live canopy. Here, percent canopy loss and percent live canopy are the response variables and the explanatory variables considered are the 20 vegetation indices from Landsat (7 pre-drought, 3 difference, and 10 post-drought), as well as the nine other variables described in section 3.3.1. Differences between post-drought and pre-drought indices were used for the three tassell cap bands to minimize correlations between the other vegetation indices.

RF is a machine learning algorithm that averages an ensemble of classification and regression trees (Breiman, 2001; Liaw & Wiener, 2002). To define the split at each node in each regression tree, only a randomly chosen subset of variables is used, in order to limit the correlation between the individual regression trees in the ensemble (Gislason, Benediktsson, & Sveinsson, 2006). RF is a non-parametric model and thus makes no assumptions as to the distributions of the underlying data. Also, RF allows for multicollinearity among variables (Piramuthu, 2008) and tends not to overfit the data (Gislason et al., 2006). Previously, RF has been used to estimate percent urban land cover (Walton, 2008) or forest cover (González-Roglich, 2016) within a pixel but only for single date images.

To model canopy loss, we used a two-step modeling approach to better account for zero-inflation (Savage et al., 2015). First we used RF in classification mode, ensemble of

classification trees, to classify pixels into two binary classes (1) absence or (2) presence of canopy loss. We then estimated the proportion of canopy loss for all pixels having some canopy loss, using RF in regression mode, ensemble of regression trees. To predict proportion post-drought live cover within a pixel, we used a one-step RF model, as an ensemble of regression trees, because there were few zeros in the post-drought live canopy dataset.

We did not find that RF was sensitive to parameterization, and therefore we selected a third of the total variables in the model to define each split. We also ran every model with 501 regression trees. For variable selection with the RF model, we first ran a full model, including all 29 explanatory variables, (Table 2), to obtain variable importance estimates. Several variable pairs were redundant ($r > 0.95$). To create a more parsimonious model, for each of the redundant pairs, one variable was removed, while preferentially keeping variables considered most important by the RF model.

The second model tested was a ZOIB regression, which is a mixture of two models: a beta distribution, which models the continuous proportions (0,1) and a degenerate distribution, which models the discrete probability mass at zero or one (Ospina & Ferrari, 2010). When both zeros and ones are present, they are modeled with a Bernoulli distribution (Ospina & Ferrari, 2012). This type of model for example, has been used to predict the percent of an area infected by mountain pine-beetle damage, using climate, slope, and elevation as explanatory variables at coarse scales (Kaufeld, Heaton, & Sain, 2014). To our knowledge, however, it has had limited use in mapping applications, such as to create continuous tree canopy loss maps using remotely sensed data.

We used Generalized Additive Models for Location, Scale and Shape (GAMLSS), (Rigby & Stasinopoulos, 2005) where the distribution for the response variable is a zero-or-one

inflated beta distribution for modeling live canopy cover or a zero-inflated beta distribution for modeling canopy loss. As an example for how canopy loss was modeled, the zero-inflated beta distribution has a probability density function as follows:

$$bi(y|\alpha, \mu, \phi) = \begin{cases} \alpha, & \text{if } y=0 \\ (1-\alpha)f(y|\mu, \phi), & \text{if } 0 < y < 1 \end{cases} \quad (\text{Ospina \& Ferrari, 2012})$$

where $f(y|\mu, \phi)$ is equal to the density function for the beta distribution, α is the probability of obtaining zero canopy loss, μ is the distribution mean (e.g. the proportion of canopy loss, given that canopy loss is greater than zero), and ϕ is a precision parameter (Ospina & Ferrari, 2012). Under the GAMLSS framework, all the parameters of the zero-inflated beta distribution (e.g. α, μ, ϕ) are modeled as additive functions of the covariates (e.g. post-drought and pre-drought Landsat vegetation indices). There is no ecological reason why certain vegetation indices would be less relevant in predicting the probability of zero canopy loss vs. the proportion of canopy loss if loss occurs. We therefore used the same covariates when modeling α and μ . Furthermore, the output predictions for these models are for each distribution parameter; as such we obtained two predictive maps of α and μ . To create the final canopy loss maps, we assumed that if the probability of obtaining zero canopy loss (α) was greater than 1/3, then the pixel had zero canopy loss; otherwise the pixel took the value of the distribution mean (μ) (e.g. the proportion of canopy loss within a pixel). We selected a probability of 1/3, because it led to the most accurate prediction of zeros in our training dataset.

For variable selection with the ZOIB regression model, a jackknifing approach was used, where (1) each explanatory variable was run in a separate model (e.g. covariate equal to one) and (2) the full model was run without that explanatory variable (e.g. total covariates minus one). These two metrics allow one to compare variable importance and variable redundancy, respectively. Variables were then ordered based on their AIC, Akaike information criterion.

Again for pairs of variables which were redundant ($r > 0.95$), the variable with the least negative AIC scores based on these two metrics was excluded.

3.3.3. Accuracy assessments of coarse-scale canopy loss maps

Model accuracy of the coarse-scale canopy loss maps was computed using two independent testing approaches. The first approach created a testing dataset that was spatially proximate to the training dataset. More specifically 33% of each orthophoto was used as testing data and 66% as training data. The second approach created testing data that was spatially independent of the training dataset. Using leave-one-out cross validation, the model was run 15 times, and for each run a single orthophoto was withheld and used as testing data, while the remaining 14 orthophotos were used as training data. After 15 model runs, each orthophoto had acted as testing data. Model accuracy was computed by averaging the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) across all 15 runs. These two accuracy assessment approaches for the spatially proximate and spatially independent testing datasets compute model performance both within and out-of-sample, respectively.

4. Results

4.1. Accuracy assessments of the fine-scale canopy loss maps

Classifications of NAIP orthophotos were highly correlated with our ground transects for both canopy loss ($R^2 = 0.90$, RMSE = 5.4%) and post-drought live canopy cover ($R^2 = 0.87$, RMSE = 11%), (Fig 3). As such, these 1-m fine-scale canopy loss maps and live canopy cover maps were used exclusively to calibrate and validate Landsat imagery, when creating regional coarse-scale canopy loss maps.

Fig. 3. Accuracy of the fine-scale canopy loss and post-drought live cover maps compared to ground measurements of dead and live canopy cover.

4.2. Comparison of accuracy amongst models

The vegetation indices most significant in each model were dependent on the response variables canopy loss and post-drought live canopy (Table 3). In general, for predicting canopy loss for either model, DVI and the tasseled cap bands were highly significant. For predicting post-drought live canopy NDVI, NBR, and the tasseled cap bands were most significant.

Table 3

Vegetation indices selected for each model: zero-inflated modeling approach using RF vs. the ZOIB regression model.

Random Forest			Zero-inflated beta regression		
Selected Vegetation Indices (ordered by variable importance)	Absence vs. presence of canopy loss	Proportion of canopy loss	Proportion of post- drought live canopy	Probability of zero canopy loss (α); proportion canopy loss (μ)	Probability of zero; probability of one; proportion post- drought live canopy
	DVI (13)	RGI (13)	NDVI (13)	DVI (13)	NDVI (13)
	TCG (13)	TCG (13)	NBR (13)	TCG (13)	NBR (13)
	RGI (13)	DVI (13)	TCG (13)	TCG (13-09)	TCW (13)
	TCB (13)	TCB (13)	TCB (13)	NDWI (13)	TCG (13)
	GNDVI (13)	NDWI (13)	TCW (13)	TCW (13-09)	RGI (13)
	(B5/B4) (13)	GNDVI (13)	RGI (13)	NDVI (13)	TCB (13)
	TCG (13-09)	TCW (13)	DVI (13)	TCB (13-09)	DVI (13)
	GNDVI (09)	TCG (13-09)		DVI (09)	
	TCW (13)	NDWI (09)		RGI (13)	
	TCW (13-09)	GNDVI (09)		TCB (13)	
	DVI (09)	TCW (13-09)		TCW (13)	
	(B5/B4) (09)	DVI (09)		RGI (09)	
	RGI (09)	TCB (13-09)		NDVI (09)	
	TCB (13-09)	RGI (09)		NDWI (09)	

13-09 = Difference between 2013 and 2009

The base model, which included only Landsat vegetation indices, was then compared to independent models, each containing one of the six sets of auxiliary variables (Table 4), for both RF and ZOIB. Error rates using the spatially proximate testing dataset were always lower compared to error rates of the spatially independent testing dataset for both RF and ZOIB. Furthermore, the percent difference between the spatially proximate and spatially independent MAE rates in predicting canopy loss were much higher for RF (45%) compared to ZOIB (14%). For predicting canopy loss, when using spatially proximate testing data, RF has the higher accuracy, whereas when using spatially independent data, ZOIB has the higher accuracy. We argue that using spatially independent testing data is the only way to understand how well the

466 model predicts canopy loss outside of where the model was parameterized, and thus ZOIB was
467 selected to predict out-of-sample results.

468

Table 4

Model comparison (60-m). Models in bold had decreasing error (MAE or RMSE) compared to the base model. These auxiliary data in bold and variables from the base model were then used in the final models to predict percent canopy loss and post-drought live canopy cover.

Cover Type	Random forest						Zero-inflated beta regression model					
	Spatially proximate			Spatially independent			Spatially proximate			Spatially independent		
	MAE	RMSE	EV*	MAE	RMSE	EV*	MAE	RMSE	AIC	MAE	RMSE	AIC
<i>Base model: Landsat spectral indices</i>												
CL	3.96	6.10	58.2	5.73	8.66	60.0	4.67	7.16	-118781	5.31	8.09	-167594
A	10.7	13.9	67.3	12.3	15.7	67.7	11.6	14.8	-55189	12.4	15.7	-77586
N	10.5	13.6		12.3	15.5		11.1	14.2		11.7	14.9	
<i>Landsat indices, % forest cover</i>												
CL	3.93	6.06	59.0	5.65	8.55	60.5	4.65	7.14	-119002	5.31	8.12	-167924
A	10.5	13.6	68.7	12.2	15.6	69.1	11.5	14.7	-55889	12.3	15.6	-78508
N	10.2	13.3		11.9	15.1		10.9	14.1		11.5	14.8	
<i>Landsat indices, latitude & longitude</i>												
CL	2.93	4.75	74.1	6.18	9.45	75.2	4.60	7.18	-120188	5.46	8.5	-169776
A	8.77	11.6	77.1	11.7	14.8	77.6	11.4	14.6	-57562	12.3	15.7	-80934
N	8.71	11.5		11.9	15.0		10.7	13.9		11.5	14.8	
<i>Landsat indices, MODIS Land Surface Temperature (LST) difference</i>												
CL	3.59	5.60	64.5	5.76	8.77	66.4	4.57	7.11	-119739	5.25	8.12	-169044
A	9.93	13.0	71.3	12.5	15.8	71.9	11.6	14.8	-55295	12.4	15.7	-77780
N	9.79	12.8		12.2	15.5		11.0	14.2		11.7	14.9	
<i>Landsat indices, elevation</i>												
CL	3.59	5.62	64.1	6.31	9.61	66.0	4.67	7.16	-118970	5.45	8.3	-167897
A	9.81	12.8	72.4	12.8	16.0	72.9	11.6	14.8	-55913	12.5	15.9	-78648
N	9.68	12.6		12.3	15.5		11.1	14.2		11.9	15.2	
<i>Landsat indices, slope & aspect</i>												
CL	3.90	5.98	60.0	5.84	8.76	61.9	4.66	7.15	-119219	5.34	8.13	-168298
A	10.2	13.2	70.1	12.3	15.6	70.8	11.4	14.5	-57157	12.4	15.7	-80438
N	9.97	13.0		12.1	15.2		10.8	13.9		11.7	14.9	
<i>Landsat indices, magnitude of change according to Landtrendr</i>												
CL	4.07	6.08	59.6	5.68	8.57	61.1	4.61	7.09	-119620	5.27	8.06	-168660
A	10.6	13.7	68.0	12.2	15.6	68.4	11.5	14.7	-55899	12.3	15.6	-78631
N	10.4	13.5		12.2	15.4		11.1	14.2		11.8	15.0	

MAE – Mean absolute error; RMSE – Root mean square error

EV – Percent variance explained for the regression model, predicting proportion canopy loss

AIC – Akaike Information Criterion

CL – Canopy Loss; A – alive; N – Non-tree

Moreover, remotely sensed derived variables (e.g. Landsat vegetation indices, percent forest cover, MODIS LST, and the magnitude of change associated with the time-series analysis) in general were more likely to improve or maintain model accuracy for both the spatially proximate and spatially independent testing datasets. In contrast, including space and landscape variables (e.g. latitude, longitude, elevation, slope, and aspect), typically decreased the error rate associated with the spatially proximate dataset but increased the error rate when validated with spatially independent testing data. However, this only occurred for models predicting canopy loss. For predicting proportion live canopy in the RF model, including latitude and longitude decreased the MAE and RMSE compared to the base model, when validated with spatially independent testing data. Furthermore, including slope and aspect slightly decreased the RMSE, and maintained the same MAE.

Final models were selected using only those explanatory variables that decreased either the MAE or RMSE, according to the spatially independent testing dataset (Table 5). Although including latitude and longitude decreased MAE for live canopy, the ranges of these variables in the training dataset were not adequately represented across the entire study area (Table 2), and therefore they were not included as covariates in the final model. The ZOIB final model had the lower error rate in predicting canopy loss (MAE = 5.16%) compared to the zero-inflated modelling approach using RF (MAE = 5.60%). Also, the ZOIB final model was capable of predicting a greater proportion of true zeros and the full range of canopy loss, compared to the RF model. Conversely, the RF model had a slightly lower but similar error rate in predicting percent live canopy cover (MAE = 12.1%) compared to the ZOIB model (MAE = 12.3%). The focus of this study was to model proportion canopy loss, and as such we elected to use the ZOIB model to predict both canopy loss and post-drought live canopy cover. We also compared speed

and memory requirements of both models in predicting canopy loss using variables from the base model. The ZOIB regression model was ~20 times faster than the RF model, and required ~5 times less memory.

Table 5
Spatially independent error rates for final canopy models comparing RF and ZOIB

Model: explanatory variables	Cover type	Observed average value, range	Predicted average value, range	Predicted # of zeros/ observed ¹	MAE	RMSE	ME
<i>Zero-inflated beta regression</i>							
Landsat indices; land surface temperature; magnitude of change	Canopy loss	(5.96%, 0-87.6%)	(5.83%, 0-91.9%)	4583/ 14582	5.16%	8.01%	+0.14%
Landsat indices; % forest cover; magnitude of change	Alive	(41.9%, 0-100%)	(41.9%, 0-100%)	-	12.3%	15.5%	-0.083%
<i>Random forest</i>							
Landsat indices; % forest cover; magnitude of change	Canopy loss	(5.96%, 0-87.6%)	(6.23%, 0-44.5%)	1929 /14582	5.60%	8.48%	-0.26%
Landsat indices; % forest cover; slope & aspect; magnitude of change	Alive	(41.9%, 0-100%)	(41.9%, 3.48%- 98.4%)	-	12.1%	15.3%	-0.070%

ME = mean error

¹The proportion of zeros in the dataset is 19%.

We also examined the distribution of post-drought live canopy and canopy loss across a canopy density gradient (pre-drought), using data from the spatially independent testing dataset. Areas with either sparse canopy or closed canopy had less canopy loss compared to intermediate-cover areas (Fig. 4).

514

515

516 **Fig. 4.** Distribution of live canopy and canopy loss across a gradient of pre-drought live canopy,
 517 from sparse to closed canopy woodlands, using data from the orthophotos. Error bars represent
 518 MAE (km²) for the canopy loss in scaling up to a 60-m grid cell.

519

520 4.3. Canopy loss estimated across the study area

521 The ZOIB regression model was used to predict post-drought live canopy cover and
 522 canopy loss using all 15 orthophotos as training data (Fig. 5). We observed an average of 4.7%
 523 canopy cover loss across all forested pixels, which is equivalent to 1,124 km² of canopy loss in
 524 our central Texas study region of ~10,850 km² of pre-drought canopy cover. Given that each
 525 pixel is not fully forested, this represents 10% loss of forested area. Predicted forest cover loss
 526 ranged from 0 to 100%.

527

Fig. 5. Coarse-scale continuous canopy loss (60-m), quantified across the study area (Landsat scene path 28/row 39) in the Llano uplift & Edwards plateau region of central Texas.

5. Discussion

We developed an approach capable of mapping drought-induced canopy loss at a regional scale that relied on high and moderate resolution remotely sensed data to create continuous canopy loss maps. Our approach works well in open canopy forests, where it is inherently more difficult to detect canopy loss, because the changes in forest cover are more subtle than in fully forested areas. As such, our dataset is composed of zero-inflated continuous proportions and we present two models capable of handling this type of data: (1) RF and (2) ZOIB.

We found the ZOIB regression model to be more robust and less computationally demanding in predicting percent tree canopy cover loss compared to RF. ZOIB had a lower error rate and was also 20 times faster, with a much lower memory requirement. When models are run using zero-inflated data, the predicted results are typically biased in two ways: (1) for areas with a zero observation, the model fails to predict the zero correctly, and instead predicts very low values, and (2) where a non-zero observation occurs, the model underestimates the value (Savage et al., 2015). The ZOIB regression model was less biased, in comparison to the zero inflated RF approach, because it predicted a greater proportion of zeros, where there were actual zeros, and the full range of canopy cover loss, where non-zero observations occurred.

The design of the training and testing datasets significantly influenced the accuracy measures. We found that it was important to validate the model with spatially independent testing data. Spatially proximate testing data would be expected to produce the highest accuracy measures as we see here, yet when the goal is to predict in new areas outside the immediate validation area, the use of spatially independent testing data is the better true assessment of model/mapping performance (González-Roglich, 2016). Furthermore, if we had only validated using spatially proximate testing data, then we would have selected the RF model as having greater accuracy. However the RF model had lower accuracy in predicting canopy loss in new areas. Therefore, during model selection, it is necessary to test models with spatially independent testing data to achieve a valid estimate of the models' power to predict out-of-sample. Assessing models with spatially proximate testing data is appropriate in cases when the model is not extended to new locations.

Variable selection was also influenced by the design of the testing and training datasets. Including topographic variables (elevation, slope, and aspect) resulted in lower error rates when

using spatially proximate testing data, but higher error rates when using spatially independent testing data for modeling canopy loss. Again if we had only validated with spatially proximate testing data, then the model would have included variables that would have decreased the models ability to predict in new areas. In general, including only remotely sensed variables led to a better predictive model for canopy loss. Conversely, for the RF model when predicting proportion live canopy, space and landscape variables had mixed effects, maintaining a similar, but slightly lower error rate for slope and aspect, decreasing the error rate for latitude and longitude, or increasing the error rate for elevation, when validated with spatially independent testing data. It is interesting that latitude and longitude played an important role in predicting percent live cover in this region, likely due to the strong east-west precipitation gradient that exists in Texas.

Studies of forest disturbance typically take place in areas where pixels are almost entirely forested. However, climate models predict that subtropical dry zones will become even drier as climate change advances (Seager et al., 2010), and these areas are often dominated by shrublands, woodlands, and savannas, which are typically open-canopy systems. Detecting mortality in these systems is a key area that merits further exploration. It is challenging to map mortality for mixed and sparsely forested pixels (compared to closed-canopy forests); however, our method of defining percent canopy loss within a pixel quantifies mortality more accurately than methods that label entire pixels as mortality. In these open canopy ecosystems, the vegetation in a 60-m grid cell is rarely completely dead, let alone entirely forested.

Drought-induced tree mortality is a subtle land cover change; however it is also widespread and can cause substantial changes in forest cover over large regions. Using the ZOIB regression model, we found an average of 4.7% canopy loss (e.g. ~10% loss of forested area). Most of this mortality occurred in areas of medium canopy cover, compared to areas with sparse

or closed-canopy (Fig. 4). Spatial patterns of drought-induced tree mortality are often complex, due to interactions between multiple controlling factors (Allen et al., 2010). For example, sites with high tree density are often more prone to density-dependent mortality (Fensham, Fairfax, & Ward, 2009), due to competition for limited water resources during a drought. However, sites with high tree density and adequate moisture can be favorable for tree growth, relative to dry landscape positions having less dense vegetation. Perhaps there is a trade-off between density-dependent factors and landscape position that could explain why tree mortality in our study region is predominantly occurring in areas with medium levels of canopy cover.

In the orthophoto classifications, the fine-scale 1-m spatial resolution allowed for the delineation of thinning/barely alive canopy and actual canopy loss. However, in scaling up to Landsat, perhaps some of the canopy loss, which we assume is due to mortality of trees, may in fact be due to canopy thinning. This could account for some of the error observed between the coarse-scale canopy loss maps created using Landsat and the fine-scale canopy loss maps created using the orthophotos.

To avoid quantifying canopy loss due to fire, we removed all areas that were burned; however, other disturbances, particularly due to insects and pathogens, could have also contributed to the canopy loss observed. Canopy loss occurred throughout the study area without definitive epi-centers inherent in most insect and disease outbreaks; therefore we assume that much of this mortality was due to drought. Nonetheless, drought-stressed oaks are more susceptible to Hypoxylon canker, which can occur in a latent state in healthy oaks. This may explain why disease incidence can increase rapidly following drought-stress (Basset, 1984). Also, oak trees already infected with oak wilt (*Ceratocystis fagacearum*) typically have weakened vascular systems, and, as such, drought can exacerbate the disease (Wilson, 2001).

Therefore, drought likely increased the susceptibility and/or accelerated the mortality risk of trees infected by insects and/or pathogens, and as such we still consider drought as a factor in killing these infected trees.

We intend to extend this analysis in the future to create canopy cover loss maps across the entire state of Texas. Texas has a significant east-west precipitation gradient and, thus, a diverse range of natural systems. Continuous maps of canopy loss across Texas will provide valuable information on how trees respond to drought stress across this large climatic gradient.

6. Conclusion

In 2011, the most severe single year drought on record caused significant tree mortality across the Edwards Plateau and Llano Uplift region of Texas. Droughts like this in Texas and elsewhere are expected to become more common as temperatures increase; reliable and robust methods are needed that can systematically detect drought-induced canopy loss. We present a semi-automated method to quantify the canopy loss resulting from this disturbance using multiple scales of publically available data, from high to moderate resolution imagery. We propose using zero-inflated beta regression models instead of linear regressions and RF models to predict percent canopy loss within a pixel, as they are more suitable for zero-inflated continuous proportion data. We detect drought-induced tree canopy loss across large regions in open/closed canopy forests, precisely the types of ecosystems (e.g. semiarid subtropical dry zones) that may see increased intensity of droughts due to climate change. Knowing where canopy loss occurs will help us to quantify ecological impacts, to calibrate and/or validate mortality algorithms in regional models, and to better understand the spatiotemporal dynamics of tree mortality.

Acknowledgements

Funding was provided by a NASA Earth and Space Science Fellowship (NNX13AN86H) and a James B. Duke Fellowship, awarded to Amanda M. Schwantes, as well as a United States Department of Agriculture/National Institute of Food and Agriculture (USDA/NIFA) grant (2012-68002-19795). We also thank Kelly Howell for her assistance in classifying NAIP images, Raven Bier & Evan Williams for field assistance, the Texas State Parks & the National Forest Service for providing access to public lands for field sampling, and Danica Schaffer-Smith & Mariano González-Roglich for offering valuable feedback.

References

- Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., ... Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, 259(4), 660–684. <http://doi.org/10.1016/j.foreco.2009.09.001>
- Allen, C. D., Breshears, D. D., & McDowell, N. G. (2015). On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere*, 6(8), 1-55. <http://doi.org/10.1890/ES15-00203.1>
- Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, 5(5), 423–431. <http://doi.org/10.1080/2150704X.2014.915434>
- Bassett, E. N., & Fenn, P. (1984). Latent colonization and pathogenicity of *Hypoxylon atropunctatum* on oaks. *Plant disease*, 68, 317-319.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <http://doi.org/10.1023/A:1010933404324>
- Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., ... Meyer, C. W. (2005). Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences of the United States of America*, 102(42), 15144–15148. <http://doi.org/10.1073/pnas.0505734102>

- 660 Clark, J. S. (1990). Landscape interactions among nitrogen mineralization, species composition,
661 and long-term fire frequency. *Biogeochemistry*, 11(1), 1–22.
- 662 Clifford, M. J., Royer, P. D., Cobb, N. S., Breshears, D. D., & Ford, P. L. (2013). Precipitation
663 thresholds and drought-induced tree die-off: Insights from patterns of *Pinus edulis* mortality
664 along an environmental stress gradient. *New Phytologist*, 200(2), 413–421.
665 <http://doi.org/10.1111/nph.12362>
- 666 Cohen, W. B., Fiorella, M., Gray, J., Helmer, E., & Anderson, K. (1998). An Efficient and
667 Accurate Method for Mapping Forest Clearcuts in the Pacific Northwest Using Landsat
668 Imagery, 64(4), 293–300.
- 669 Cohen, W. B., & Goward, S. N. (2004). Landsat's Role in Ecological Applications of Remote
670 Sensing. *BioScience*, 54(6), 535–545. [http://doi.org/10.1641/0006-](http://doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2)
671 [3568\(2004\)054\[0535:LRIEAO\]2.0.CO;2](http://doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2)
- 672 Collins, B. M., Miller, J. D., Thode, A. E., Kelly, M., van Wagendonk, J. W., & Stephens, S. L.
673 (2009). Interactions Among Wildland Fires in a Long-Established Sierra Nevada Natural
674 Fire Area. *Ecosystems*, 12(1), 114–128. <http://doi.org/10.1007/s10021-008-9211-7>
- 675 Coops, N. C., Johnson, M., Wulder, M. a., & White, J. C. (2006). Assessment of QuickBird high
676 spatial resolution imagery to detect red attack damage due to mountain pine beetle
677 infestation. *Remote Sensing of Environment*, 103(1), 67–80.
678 <http://doi.org/10.1016/j.rse.2006.03.012>
- 679 Crist, E. P. (1985). A TM Tasseled Cap equivalent transformation for reflectance factor data.
680 *Remote Sensing of Environment*, 17, 301–306. [http://doi.org/10.1016/0034-4257\(85\)90102-](http://doi.org/10.1016/0034-4257(85)90102-6)
681 [6](http://doi.org/10.1016/0034-4257(85)90102-6)
- 682 Elliott, L. F., Diamond, D. D., True, C. D., Blodgett, C. F., Pursell, D., German, D., & Treuer-
683 Kuehn, A. (2014). Ecological Mapping Systems of Texas: Summary Report. Austin: Texas
684 Parks & Wildlife Department. Available at: <http://tpwd.texas.gov/gis/gallery/> [February
685 2014].
- 686 Engelbrecht, B. M. J., Comita, L. S., Condit, R., Kursar, T. a, Tyree, M. T., Turner, B. L., &
687 Hubbell, S. P. (2007). Drought sensitivity shapes species distribution patterns in tropical
688 forests. *Nature*, 447(7140), 80–82. <http://doi.org/10.1038/nature05747>
- 689 Fensham, R. J., Fairfax, R. J., & Ward, D. P. (2009). Drought-induced tree death in savanna.
690 *Global Change Biology*, 15(2), 380–387. <http://doi.org/10.1111/j.1365-2486.2008.01718.x>
- 691 Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover
692 classification. *Pattern Recognition Letters*, 27(4), 294–300.
693 <http://doi.org/10.1016/j.patrec.2005.08.011>

- 694 Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote
695 sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3),
696 289–298. [http://doi.org/10.1016/S0034-4257\(96\)00072-7](http://doi.org/10.1016/S0034-4257(96)00072-7)
- 697 González-Roglich, M., & Swenson, J. J. (2016). Tree cover and carbon mapping of Argentine
698 savannas: Scaling from field to region. *Remote Sensing of Environment*, 172, 139–147.
699 <http://doi.org/10.1016/j.rse.2015.11.021>
- 700 Guardiola-Claramonte, M., Troch, P. A., Breshears, D. D., Huxman, T. E., Switanek, M. B.,
701 Durcik, M., & Cobb, N. S. (2011). Decreased streamflow in semi-arid basins following
702 drought-induced tree die-off: A counter-intuitive and indirect climate impact on hydrology.
703 *Journal of Hydrology*, 406(3), 225–233. <http://doi.org/10.1016/j.jhydrol.2011.06.017>
- 704 Hoerling, M., Kumar, A., Dole, R., Nielsen-Gammon, J. W., Eischeid, J., Perlwitz, J., ... Chen,
705 M. (2013). Anatomy of an extreme event. *Journal of Climate*, 26(9), 2811–2832.
706 <http://doi.org/10.1175/JCLI-D-12-00270.1>
- 707 Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., ... Megown, K. (2015).
708 Completion of the 2011 National Land Cover Database for the conterminous United States-
709 Representing a decade of land cover change information. *Photogrammetric Engineering*
710 *and Remote Sensing*, 81(5), 345–354.
- 711 Kaufeld, K. A., Heaton, M. J., & Sain, S. R. (2014). A Spatio-Temporal Model for Mountain
712 Pine Beetle Damage. *Journal of Agricultural, Biological, and Environmental Statistics*,
713 19(4), 439–452. <http://doi.org/10.1007/s13253-014-0182-1>
- 714 Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and
715 recovery using yearly Landsat time series: 1. LandTrendr - Temporal segmentation
716 algorithms. *Remote Sensing of Environment*, 114(12), 2897–2910.
717 <http://doi.org/10.1016/j.rse.2010.07.008>
- 718 Kieschnick, R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1):
719 percentages, proportions and fractions. *Statistical modelling*, 3(3), 193–213.
720 <http://doi.org/10.1191/1471082X03st053oa>
- 721 Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3),
722 18–22.
- 723 McDowell, N. G., Beerling, D. J., Breshears, D. D., Fisher, R. A., Raffa, K. F., & Stitt, M.
724 (2011). The interdependence of mechanisms underlying climate-driven vegetation
725 mortality. *Trends in Ecology and Evolution*, 26(10), 523–532.
726 <http://doi.org/10.1016/j.tree.2011.06.003>
- 727 McDowell, N. G., Coops, N. C., Beck, P. S. A., Chambers, J. Q., Gangodagamage, C., Hicke, J.
728 a., ... Allen, C. D. (2015). Global satellite monitoring of climate-induced vegetation

- disturbances. *Trends in Plant Science*, 20(2), 114–123.
<http://doi.org/10.1016/j.tplants.2014.10.008>
- Meddens, A. J. H., & Hicke, J. A. (2014). Spatial and temporal patterns of Landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. *Forest Ecology and Management*, 322, 78–88. <http://doi.org/10.1016/j.foreco.2014.02.037>
- Meigs, G. W., Kennedy, R. E., & Cohen, W. B. (2011). A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. *Remote Sensing of Environment*, 115(12), 3707–3718.
<http://doi.org/10.1016/j.rse.2011.09.009>
- Michaelian, M., Hogg, E. H., Hall, R. J., & Arsenault, E. (2011). Massive mortality of aspen following severe drought along the southern edge of the Canadian boreal forest. *Global Change Biology*, 17(6), 2084–2094. <http://doi.org/10.1111/j.1365-2486.2010.02357.x>
- Moore, G. W., Edgar, C. B., Washington-Allen, R. A., Vogel, J. G., March, R. G., & Zehnder, R. (In Press) Tree mortality from an exceptional drought spanning mesic to semiarid ecoregions. *Ecological Applications*. <http://dx.doi.org/10.1890/15-0330.1>
- Mueller, R. C., Scudder, C. M., Porter, M. E., Talbot Trotter, R., Gehring, C. A., & Whitham, T. G. (2005). Differential tree mortality in response to severe drought: Evidence for long-term vegetation shifts. *Journal of Ecology*, 93(6), 1085–1093. <http://doi.org/10.1111/j.1365-2745.2005.01042.x>
- Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, 51(1), 111–126. <http://doi.org/10.1007/s00362-008-0125-4>
- Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609–1623.
<http://doi.org/10.1016/j.csda.2011.10.005>
- Piramuthu, S. (2008). Input data for decision trees. *Expert Systems with Applications*, 34(2), 1220–1226. <http://doi.org/10.1016/j.eswa.2006.12.030>
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554. <http://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Rogan, J., Franklin, J., & Roberts, D. A. (2002). A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80(1), 143–156. [http://doi.org/10.1016/S0034-4257\(01\)00296-6](http://doi.org/10.1016/S0034-4257(01)00296-6)
- Rupp, D. E., Li, S., Massey, N., Sparrow, S. N., Mote, P. W., & Allen, M. (2015). Anthropogenic influence on the changing likelihood of an exceptionally warm summer in

763 Texas, 2011. *Geophysical Research Letters*, 42(7), 2392–2400.
 764 <http://doi.org/10.1002/2014GL062683>

765 Savage, S. L., Lawrence, R. L., & Squires, J. R. (2015). Predicting relative species composition
 766 within mixed conifer forest pixels using zero-inflated models and Landsat imagery. *Remote*
 767 *Sensing of Environment*, 171, 326–336. <http://doi.org/10.1016/j.rse.2015.10.013>

768 Seager, R., Naik, N., & Vecchi, G. A. (2010). Thermodynamic and dynamic mechanisms for
 769 large-scale changes in the hydrological cycle in response to global warming. *Journal of*
 770 *Climate*, 23(17), 4651–4668. <http://doi.org/10.1175/2010JCLI3655.1>

771 Thornton, P. E., Running, S. W., & White, M. A. (1997). Generating surfaces of daily
 772 meteorological variables over large regions of complex terrain. *Journal of Hydrology*, 190,
 773 214–251. [http://doi.org/10.1016/S0022-1694\(96\)03128-9](http://doi.org/10.1016/S0022-1694(96)03128-9)

774 Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., &
 775 Cook, R. B. (2014). DAYMET: Daily Surface Weather on a 1 km Grid for North America,
 776 Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. Available at:
 777 <http://daymet.ornl.gov/>. [June 2015]. <http://dx.doi.org/10.3334/ORNLDAAC/1219>

778 Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., &
 779 Sheffield, J. (2014). Global warming and changes in drought. *Nature Clim. Change*, 4(1),
 780 17–22. <http://doi.org/10.1038/NCLIMATE2067>

781 Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring
 782 vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [http://doi.org/10.1016/0034-](http://doi.org/10.1016/0034-4257(79)90013-0)
 783 [4257\(79\)90013-0](http://doi.org/10.1016/0034-4257(79)90013-0)

784 US Department of Agriculture. (2014). Geospatial data gateway. Available at:
 785 <http://datagateway.nrcs.usda.gov/>. [March 2014].

786 USDA Forest Service/U.S. Geological Survey. (2014). MTBS Data Access: Fire Level
 787 Geospatial Data. Available at: <http://mtbs.gov/dataquery/individualfiredata.html>. [Nov
 788 2014].

789 USDA Natural Resources Conservation Service. (2015). The PLANTS Database. National Plant
 790 Data Team, Greensboro, NC, USA. Available at: <http://plants.usda.gov>.

791 United States Geological Survey. (2015). Product guide: Landsat 4-7 Climate data record (CDR)
 792 surface reflectance. Version 6.1.

793 Van Mantgem, P. J., Stephenson, N. L., Byrne, J. C., Daniels, L. D., Franklin, J. F., Fulé, P. Z.,
 794 ... Veblen, T. T. (2009). Widespread increase of tree mortality rates in the western United
 795 States. *Science*, 323(5913), 521–524. <http://doi.org/10.1126/science.1165000>

- 796 Van Wagtendonk, J. W., Root, R. R., & Key, C. H. (2004). Comparison of AVIRIS and Landsat
797 ETM+ detection capabilities for burn severity. *Remote Sensing of Environment*, 92(3), 397–
798 408. <http://doi.org/10.1016/j.rse.2003.12.015>
- 799 Vogelmann, J. E., & Rock, B. N. (1988). Assessing forest damage in high-elevation coniferous
800 forests in Vermont and New Hampshire using thematic mapper data. *Remote Sensing of*
801 *Environment*, 24(2), 227–246. [http://doi.org/10.1016/0034-4257\(88\)90027-2](http://doi.org/10.1016/0034-4257(88)90027-2)
- 802 Walton, J. T. (2008). Subpixel Urban Land Cover Estimation: Comparing Cubist, Random
803 Forests and Support Vector Regression. *Photogrammetric Engineering & Remote Sensing*,
804 74(10), 1213–1222.
- 805 Wan, Z. (2008). New refinements and validation of the MODIS Land-Surface
806 Temperature/Emissivity products. *Remote Sensing of Environment*, 112(1), 59–74.
807 <http://doi.org/10.1016/j.rse.2006.06.026>
- 808 Willson, C. J., Manos, P. S., & Jackson, R. B. (2008). Hydraulic traits are influenced by
809 phylogenetic history in the drought-resistant, invasive genus *Juniperus* (Cupressaceae).
810 *American Journal of Botany*, 95(3), 299–314. <http://doi.org/10.3732/ajb.95.3.299>
- 811 Wilson, A. D. (2001). Oak wilt: a potential threat to southern and western oak forests. *Journal of*
812 *Forestry*, 99(5), 4–11.
- 813 Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of
814 Landsat TM imagery. *Remote Sensing of Environment*, 80(3), 385–396.
815 [http://doi.org/10.1016/S0034-4257\(01\)00318-2](http://doi.org/10.1016/S0034-4257(01)00318-2)

816

Graphical abstract