**Next-generation mapping for regional smoke management and emissions inventories: incorporating underlying uncertainty in wildland fuel characterization**

Prichard, S.J., Kennedy, M.C., Andreu, A., Eagle, P.E., and French, N.

**Introduction**

Raster maps of vegetation and biomass are increasingly used in wildfire hazard assessments (Rollins 2009, Scott et al. 2013), emissions inventories (e.g., US EPA 2017), carbon mapping (Blackard et al. 2007, Pan et al. 2011) and local to regional wildland fire and smoke management planning. Traditionally, single biomass values have been assigned to mapped pixels, often based on broadly classified vegetation type and assignment using look-up tables or nearest neighbor imputation methods (e.g., Rollins et al. 2004, Pierce et al. 2009, Keane et al. 2013, Riley et al. 2016). Ideally, maps of vegetation and combustible biomass (wildland fuels) would be based on highly replicated and accurate field measurements to calibrate relationships with remotely sensed imagery and include estimated uncertainty. However, wildland fuels are highly dynamic and variable across time and space (Keane et al. 2012), and their inherent variability generally makes it impractical to collect enough measurements to represent the broad geographic and structural diversity of wildland vegetation and fuels.

Fuel mapping therefore generally relies on classifications of fuels based on mapped vegetation or interpretation of remotely sensed imagery rather than measured values (Keane et al. 2015). For example, LANDFIRE maps surface and canopy fuel characteristics across the United States, assigning fuelbeds from the Fuel Characteristics Classification System (Ottmar et al. 2007), fuel loading models (Lutes et al. 2009), fire behavior fuel models (Anderson 1982) and canopy characteristics based on remotely sensed imagery and other data layers (Rollins et al. 2009). Mapped fuels are often represented as estimates of biomass per area by major vegetation type, and some representations also include estimates of biomass by fuel stratum category (e.g., canopy, shrub, herbaceous, downed wood by size class, litter and organic soil or duff). For biomass or emissions inventories (e.g., US EPA 2017) fuel characteristics are summarized at relatively coarse scales (1-km pixels) and aggregate finer-scale variability in vegetation.

Remotely sensed datasets are generally useful for characterizing upper canopy fuels of dense forests and shrublands but are not suitable for surface fuel characterization. Because surface fuels are highly variable and are not easily predicted from canopy information, there are generally high misclassification rates when canopy fuels data are used to predict surface fuels (REFS). In addition, the majority of combustible biomass of surface fuels is within the first 10-cm of the fuelbed and often composed of intermixed live and dead vegetation including shrubs, grasses, downed wood and litter that is difficult to quantify from remotely sensed imagery (Rowell XXXX).

Uncertainty in biomass estimates underlies any existing fuel classification and is generally not acknowledged, much less quantified (Congalton et al. 2014; Urbanski et al. 2011). In general, regional assessments use point-based estimates as inputs to model applications and to inform management decisions. However, whether the biomass is empirically estimated or modeled, the biomass assigned to each mapped pixel has inherent sampling or prediction error. To avoid false precision in biomass estimates and resulting model predictions, it is imperative to inform uncertainty bounds on mapped biomass values and how uncertainty is propagated within predictive models and may potentially influence management decisions. Quantifying the uncertainty in mapped biomass values has important consequences for how the biomass values are used downstream in model applications. For example, with sufficient replication, field-based inventories of site biomass can be used to quantify a prediction interval for total biomass. For model applications, there are multiple sources that contribute to the total uncertainty in model prediction, and these can be classified broadly into four groups: structure uncertainty, parameter estimation uncertainty, data input uncertainty, and natural variability/stochasticity (O’Neill and Gardner, 1979; Beck, 1987; Turley and Ford, 2009). For model applications that require fuel loadings as data input, it is important to understand how uncertainty in fuel loading values propagate to uncertainty in model predictions (Hanna 1988). Characterizing a pixel as a single point estimate masks the underlying uncertainty, resulting in false precision in model predictions. A credible estimate of the uncertainty (or variability) in fuel loading associated with a given pixel or location requires knowledge of the likely distribution of the input data.

When we consider the quality of current continental-scale fuel maps, it is not particularly informative to attempt to validate the fuel loading of an individual pixel against plot-level data. Plot-level data are subject to sampling variability and may not represent the full pixel—such a validation will inevitably fail. Although point estimates do represent our best single value for an individual pixel, a point estimate individually will not match exactly the true on-the ground value. Fuels are variable at multiple spatial scales, and to improve the quality of our mapped products we must quantify this variability to understand uncertainty in fuel loading underlying each mapped point value. This has two main contributions: 1) guide sampling efforts to improve uncertainty bounds for vegetation types for which the uncertainty bounds are wide and/or data are insufficient to estimate distributions of loadings, and 2) provide distributions for data input to modeling applications to facilitate uncertainty analysis in model prediction and associated decision-making.

The primary goal of this study was provide the capacity to characterize the distributions underlying mapped fuel inputs in order to facilitate uncertainty analysis in modeling applications. To do this, we developed a geospatial database of quantified fuel loading values to characterize the inherent variability of fuels within and across major vegetation types of the United States and Canada and to identify gaps in fuels observations. For vegetation types that had sufficient quantification of fuels by major category (e.g., canopy, shrub, herbaceous, downed wood, litter and duff), we developed empirical distribution estimates of observed fuel biomass, hereafter referred to as fuel loads, by major category. Published distributions will be useful for informing the first-generation fuels mapping that incorporates uncertainty estimates by major fuel category. Results of this study also will help inform future sampling needs to better represent the biomass of wildland fuels. Because fuel loads are a common input in fire and smoke models, uncertainty intervals informed by the fuel loading database can be used to better understand uncertainty in predictions of wildland fuel consumption and emissions and in regional to national mapping applications for biomass, carbon and emissions inventories.

**Methods**

*Fuel loading database*

The U.S. Fuel Loading Database was created as part of a JFSP-funded project (15-1-01-1 Mapping Fuels for Regional Smoke Management and Emissions Inventories). The database stores existing dry-weight biomass observations by major fuel category (i.e., tree crowns, snags, shrubs, herb, downed wood by size class, litter and duff) across the United States. Our team began by compiling existing databases and importing fuel loadings in a standard unit of measure (Mg/ha). Existing databases, including the source data for fuel loading models (Lutes et al. 2009) and LANDFIRE public source reference database (https://www.landfire.gov/lfrdb.php) were compilations of published literature and plot data. Table 1 lists the databases and provides a brief description and a source reference. We next conducted a literature review of biomass, fuel characterization and fuel consumption literature and added over 150 individual references. Minimum standards for including observations in the database were that they: 1) contained a source reference such as FIA inventory plot and sample year or journal article citation, 2) had an identifiable vegetation type, and 3) relied on field measurements as opposed to photo monitoring sites or other visual estimations.

To group fuel loading observations by vegetation type, a standard mapping classification was needed. Because LANDFIRE is a widely used mapping source of geospatial fuels and vegetation, we chose to use Existing Vegetation Type Group (www.landfire.gov/NationalProductDescriptions21.php). There are 640 existing vegetation types within LANDFIRE and a total of 207 EVT Groups. Given that the objective of the database was to quantify the distribution of fuel loads within vegetation types, we opted to use a more generic vegetation classification (EVT Group), which is provided within the LANDFIRE EVT layer, to ensure greater numbers of records within each vegetation group. It also reduced uncertainty in assigning vegetation type to each record. Most records within the database had either a general description of vegetation, a listing of major species, a Society of American Foresters or Society of Rangeland Management cover type, or a more general Forest Type (e.g., FIA plots). We developed crosswalk tables to convert cover and forest types to EVT Groups. For records that only had a general vegetation description, we individually assigned a vegetation type.

As the database was assembled, we performed a series of quality assurance and control measures. We first screened any records that did not have geospatial location. For each of these records, we attempted to assign a geospatial location and standardized existing location data into latitude and longitude (decimal degrees). In some cases, it was necessary to assign site locations based on site descriptions. Many records (n = 2470) had geospatial location but no associated vegetation type or information. For these, we overlaid record locations with the EVT Groups layer in ArcGIS and assigned a likely EVT Group based on spatial location. Due to the potential errors incurred by spatial assignment, we tagged each of these records as having spatially-assigned vegetation types. Fuel loading values were summarized into fields defined in Table 2. In many instances, simple summations were required to create summary inputs (e.g. herb load was calculated as the sum of forb and graminoid loadings and total CWD is the sum of all sound and rotten coarse wood classes).

The fuel loading database includes data from 292 sources from existing fuel loading databases and scientific literature. Entries from existing databases were presumed to be quality checked by the source agency and were not rechecked. A random selection of 15-20% of the literature sources were checked for data entry errors, including errors in unit conversions, standardizations of woody fuel size classes, and site descriptors. Due to the extent of data entry errors an additional 10% of the literature sources were checked to ensure a higher level of accuracy. *Should we report any error rate?*

For every record that had a published source reference, we obtained the source reference and included a full citation. For quality assurance and quality control, we subsampled 30% of all source references and confirmed that entered data was accurate. Most identified errors were simple rounding errors and were corrected. In a few cases, some fuel categories were missing from the inputs and were added from the published source. In other cases, fuel categories were inaccurate and corrected within the database entries.

As the database was compiled, supported fuel loading fields were expanded to accommodate various studies and approaches. Table 2 presents the fuel categories and definitions within the database. Many categories are sparsely populated but are included because they are important within particular EVT Groups. For example, moss and ground lichen are important in many boreal and subboreal vegetation types but are relatively rare in other ecosystems and associated EVT Groups.

**Table 1**: Major source databases within the Fuel Loading database.

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | **Number of records** | **Years** | **Source** |
| FFS | 128 |  | Fire and fire surrogates (McIvor) |
| FLM database | 8555 |  | Source data for the fuel loading model development (Keane) |
| FOFEM fuels | 1095 |  | Old database compiled to inform FOFEM fuel loading profiles (Reinhardt, Lutes) |
| Forest Inventory and Analysis Program | 15,061 | 2015 | David Chojnacky, University of Vermont– downloaded from - <http://web.gis.vt.edu/forestry/dwm/index.php> |
| LFRDB | 18,012 |  | LFRDB\_Public\_20100122.mdb |
| Natural Fuels Photo Series | 550 | 1998-2016 | <https://www.fs.fed.us/pnw/fera/research/fuels/photo_series> |

**Table 2**: Fuel loading database fields and definitions. To date, the database contains nearly 40,000 records and was designed to accommodate additional records as they become available.

|  |  |  |
| --- | --- | --- |
| **Field** | **Definition** | **Sample entry** |
| LFEVTGroupID | Unique ID for each EVT Group number | 693 |
| LFEVTGroup | EVT Group Name | Spruce-Fir-Hardwood Forest |
| sourceID | Unique ID for each source reference | 571 |
| Source | Source reference | Natural Fuels Photo Series Volume Iia, PMS 836 |
| studyPointID | Unique study point ID | 48753 |
| Plotname | Plot name if provided | AKHD 15 |
| State | State name | AK |
| inventoryYear | Inventory or sampling year | 2007 |
| veg\_notes | Vegetation description | Closed spruce-paper birch forest |
| us\_loading: Mg/ha | Understory crown loading (check) | 1.52 |
| ms\_loading: Mg/ha | Midstory crown loading (check) | 22.88 |
| os\_loading: Mg/ha | Overstory crown loading (check) | 91.32 |
| tree\_crown\_loading: Mg/ha | Total tree crown loading - sum of understory, midstory and overstory |  |
| tree\_loading: Mg/ha | Total aboveground tree biomass, including boles |  |
| snag\_loading: Mg/ha | Total aboveground biomass of dead trees, all decay classes | 13.56 |
| shrub\_loading: Mg/ha | Total aboveground biomass of shrubs | 3.43 |
| herb\_loading: Mg/ha | Total aboveground biomass of herbaceous plants including grasses and other nonwoody plants | 0.06 |
| 1hr\_loading: Mg/ha | 0-1/4 inch or 0.67 cm diameter wood | 0.9 |
| 10hr\_loading: Mg/ha | 1/4 to 1 inch or 0.67 to 2.54 cm diameter wood | 1.34 |
| 100hr\_loading: Mg/ha | 1-3 inch or 2.54 to 7.6 cm diameter wood | 2.46 |
| fwd\_loading: Mg/ha | Sum of fine wood (1, 10, 100-hr) wood |  |
| 1KhrS\_loading: Mg/ha | Sound wood 3 to 9 inches or 7.62 to 22.86 cm diameter (S1000hr wood) | 0.22 |
| 1KhrR\_loading: Mg/ha | Rotten wood 3 to 9 inches or 7.62 to 22.86 cm diameter (R1000hr wood) | 0 |
| 1Khr\_loading: Mg/ha | Sum of 1000hr wood |  |
| 10KhrS\_loading: Mg/ha | Sound wood 9 to 20 inches or 22.86 to 50.8 cm diameter (S10,000hr wood) | 0 |
| 10KhrR\_loading: Mg/ha | Rotten wood 9 to 20 inches or 22.86 to 50.8 cm diameter (R10,000hr wood) | 0 |
| 10Khr\_loading: Mg/ha | Sum of 10,000hr wood |  |
| GT10KhrS\_loading: Mg/ha | Sound wood > 20 inches or 50.8 cm diameter (S >10,000hr wood) |  |
| GT10KhrR\_loading: Mg/ha | Rotten wood > 20 inches or 50.8 cm diameter (R >10,000hr wood) |  |
| GT10Khr\_loading: Mg/ha | Sum of >10,000hr wood |  |
| cwd\_sound\_loading: Mg/ha | Sum of sound coarse wood (1000, 10,000, and >10,000hr wood) |  |
| cwd\_rotten\_loading: Mg/ha | Sum of rotten coarse wood (1000, 10,000, and >10,000hr wood) |  |
| cwd\_loading: Mg/ha | Sum of coarse wood (1000, 10,000, and >10,000hr wood) |  |
| moss\_loading: Mg/ha | Biomass of surface fuel cryptograms (arboreal moss not included) | 1.48 |
| lichen\_loading: Mg/ha | Biomass of ground lichens (arboreal lichens not included) | 0 |
| litter\_depth: cm | Depth of the litter layer (Oi soil layer) is included because many sources record this instead of loading. A generic bulk density value can be used to estimate biomass from this. |  |
| litter\_loading: Mg/ha | Litter biomass (Oi soil layer) | 4.68 |
| duff\_depth: cm | Depth of the duff layer (Oe and Oa soil layers) is included because many sources record this instead of loading. A generic bulk density value can be used to estimate biomass from this. |  |
| duff\_loading: Mg/ha | Duff biomass (combined upper and lower duff layers) |  |

*Fuel loading distributions*

Database values were clustered by LANDFIRE EVT Groups for estimation of loading distributions. All analyses were conducted in the R statistical program (version 3.4.1; (R Core Team 2017)), and distributions estimated using the R fitdistr package (Delignette-Muller and Dutang 2015). To identify candidate distributions for individual fuel loading categories, an exploratory data analysis (EDA) was conducted on select EVT Groups with substantial representation (> 1000 entries). Histograms, boxplots, and normal QQ plots were used to understand prominent distribution shapes and to assist in QA/QC of the database. This exploratory data analysis showed that many of the fuel types had a high proportion of values that were zero, and the fuel loading distributions tended to be right-skewed rather than symmetric. Due to these features, we chose a hurdle estimation procedure, described in the next section.

*Hurdle distribution fitting*

It is common in empirical studies of biomass (or, more commonly, abundance) for there to be excessive density at zero (Welsh et al. 1996, Lecomte et al. 2013) relative to the density functions commonly estimated for such data. Often the non-zero distribution is skewed to the right, implying a distribution such as the log-normal or the gamma distribution is more appropriate than the normal distribution (Lecomte et al. 2013). One method to contend with excessive density at zero is to estimate two models for the data, one that predicts the probability of observing a zero, and a second that models the distribution of non-zero values (Welsh et al. 1996, Lachenbruch 2002). It can be shown that the maximum likelihood estimate for the two-part model can be obtained by finding maximum likelihood estimates for each part individually (Duan et al. 1983, Welsh et al. 1996). Such a two-stage (two-part) estimation procedure has been called by many names, but we will use the nomenclature of a "hurdle model." Qualitatively, the hurdle to be crossed is having a non-zero fuel loading, and once that hurdle is crossed (x>0) a continuous distribution is estimated for the data. The density function for the jth fuel type in the kth EVT group (fkj(x)) can be written as (Lachenbruch 2002):

fkj(x,d)= πkj1-d ((1-πkj)hkj(x))d, (1)

where h(x) is the estimated continuous distribution function (in this case, gamma or lognormal) for x>0, d = 1 if x non-zero, 0 if x 0, and π is the probability of observing a zero. For this distribution, the expected value is:

E(x) = (1-π)E(h(x)) (2)

For the continuous portion of each fuel type in each EVT group we estimated and compared lognormal and gamma distributions. The lognormal probability distribution function, with parameters μ, σ, is written as:

x > 0 (3),

where σ is the standard deviation of ln (x) and μ is the mean of ln(x). The expected value of the lognormal distribution is:

(4)

The gamma probability distribution function, with parameters α, β, is written as:

x>0 (5)

With expected value:

E(x) = αβ (6)

Estimation of the hurdle distribution occurs in two steps. Let nkj be the total number of entries in the database for a particular fuel type (j) in a particular EVT group (k), and xkji be the ith fuel loading value for fuel type j in EVT group k. Then:

1. Estimate where I is an indicator function that takes a value of 1 if the entry has a value of 0, 0 otherwise and is the estimation probability of zero loading.

2. For the remaining non-zero entries (x), use the fitdistr function in the R fitdistrplus package to find the maximum likelihood estimates of distribution parameters for the lognormal and gamma distributions.

For initial distribution fitting we decided on a minimum of 30 *non-zero entries* required for a distribution to be estimated. This balanced our ability to estimate more distributions with the uncertainty in estimating distributions for small sample sizes. With 95% confidence n = 30 is expected to obtain an estimated distribution with cumulative distribution function at most 0.25 away from the true cumulative distribution (Massart 1990).

Assessing distribution estimates

There are 30 total fuel types, and a total of 134 EVT groups in the current database. In general it is best practice to assess distribution fits graphically, but this is untenable with so many individual distributions to be estimated in the database. Instead, we use several goodness of fit quantities to evaluate the distribution fits.

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test is used for the null hypothesis that a given data set follows a specified theoretical distribution. In general it is designed for situations where the full theoretical distribution is specified *a priori* and performs poorly if distribution parameter values estimated from the data are used to specify the distribution for the KS test (Lilliefors 1967). We use a Monte-Carlo (MC) procedure to estimate the p-value for the estimated distribution against the data, where a smaller p-value indicates that the observed data is statistically different than the estimated distribution (following Lilliefors 1967)).

In the MC procedure, we calculate KS statistic for observed distribution relative to "theoretical" distribution at estimated parameter values. Then for 5000 MC replicates we take n (n=number of observed values in original distribution fit) random draws from the "theoretical" distribution at estimated parameter values. For each of these, we estimate the same theoretical distribution, then perform KS test of random to theoretical distribution at estimated parameter values. This generates 5000 KS values when the null hypothesis is true, thus a "null" distribution. The p-value is then calculated as:

(7)

where nmc is the number of MC replicates in the null distribution (5000), di is the value of the KS statistic for MC replicate i, and dobs is the observed value of the KS statistic. I is an indicator function that takes a value of 1 if the observed statistic is greater than the simulated, 0 otherwise. The sum tallies the number of simulated statistics are smaller than the observed statistic. We divide by nmc+1 because we have nmc+1total statistics (including dobs). We can then evaluate, against some α level of significance, which distributions are "fail to reject" (FTR).

For an application like this, interpretation of the KS test suffers from two issues related to sample size. At low sample sizes the test has insufficient statistical power to reject the null hypothesis; in these instances, result FTR does not necessarily provide evidence in favor of the estimated distribution (the null hypothesis). A large sample size presents the opposite problem: as sample size increases, the effect size necessary to reject the null hypothesis decreases. At large sample sizes this means that although the observed data are statistically different than the estimated distribution, the difference may not be of practical significance. For these reasons, we use equivalence tests to aid our interpretation of the goodness of fit between observed data and estimated distributions.

Equivalence tests

Robinson and Froese (2004) recommend an equivalence test to compare empirical data to model predictions using a two-one-sided t-test (TOST). In equivalence testing a maximum allowable error (or error tolerance) is defined, and the null hypothesis is that the observed distribution is outside of the error tolerance relative to a theoretical distribution. If the observed distribution is seen to be within the maximum error (or error tolerance), then the null hypothesis is rejected and the observed data is judged to be "equivalent" to the theoretical distribution (within the error tolerance). Here we use TOST to assess adequate matching between our observed empirical cumulative distribution of fuel type and the theoretical cumulative distribution function (CDF) associated with each candidate distribution. Let x(i) be the ith quantile of the empirical data distribution, and be the ith quantile of the theoretical distribution. Then the difference between the observed and theoretical cumulative distributions (xdi) is:

We then calculate as the mean distance between observed and theoretical cumulative distributions and use TOST to determine statistically if the observed and theoretical distributions differ by more than a specified error tolerance ε. This requires an error tolerance to be specified, which for our application would be a relatively arbitrarily defined threshold.

Prichard et al. (2014) use a similar equivalence procedure to evaluate the uncertainty of the fits of observed fuel consumption relative to those predicted by empirical consumption equations. For their analysis, rather than choosing a single arbitrary error threshold, they repeated the equivalence test with increasing ε until the first epsilon at which the equivalence test null hypothesis was rejected. This then defined the bound of uncertainty for that fuel type. We adapt their approach here, repeating the equivalence test for increasing error thresholds between observed and theoretical distributions for distributions estimated both with zeroes (and an offset), and distributions estimated for only values > 0. We then compare the minimum ε that rejects the null hypothesis to assess the uncertainty in the distribution estimates.

For assessing distribution estimates, the best fits would be fuel types with a KS p-value > α, and a small ε value for the equivalence procedure outlined above. We assigned broad goodness of fit classifications based on these two goodness of fit metrics (Table 3) in combination with the sample size. A fit was considered excellent if it was based on ≥ 100 entries, associated with a non-significant KS MC p-value, and had an ε value ≤ 0.05. A fit was considered good if it was based on ≥ 30 entries, had a non-significant KS p-value, and 0.05 < ε ≤ 0.15; alternatively, a fit was considered good if it has > 30 entries, and a significant KS MC p-value associated with an ε ≤ 0.05. A fit was considered poor if it has ≥ 30 entries, associated with a significant KS p-value and a large (> 0.15) ε value. The distribution was not estimated for any fuel type X EVT group combination with < 30 entries, and assigned an NA here.

**Table 3**: Criteria to rate quality of distribution fits. Ratings are excellent, good, or poor. Criteria are based on the number of entries in the database (higher n is evidence for a quality fit), the Kolmogorov-Smirnov Monte Carlo p-value (p>0.05 is evidence for a quality fit), and error threshold for equivalence test (ε at which the equivalence test null hypothesis is first rejected; a lower ε is evidence for a quality fit). Note that there are two criteria combinations to achieve a “Good” rating depending on the criterion by which an Excellent fit is not obtained (e.g., either the KS p-value is too small, or the ε value is too large). A poor fit is obtained when both the criteria for an Excellent fit are not met. An Excellent fit also requires n≥100.

|  |  |  |  |
| --- | --- | --- | --- |
| n | KS | ε | Rating |
| ≥100 | > 0.05 | (0,0.05] | Excellent |
|  | > 0.05 | (0.05,0.15] | Good |
|  | < 0.05 | (0, 0.05] | Good |
|  | < 0.05 | > 0.05 | Poor |
| [30,100) | > 0.05 | ≤ 0.15 | Good |
|  | < 0.05 | ≤ 0.05 | Good |
|  | < 0.05 | > 0.05 | Poor |
|  |  |  |  |

Uncertainty in distribution estimates

Finally, we use a bootstrap procedure to estimate a standard deviation for estimated distribution parameter values and to generate a 95% confidence interval for each distribution parameter value. The bootstrap estimates are generated using the bootdist function in the fitdistrplus package in R. In general the observed data are resampled with replacement and the distribution parameters estimated for each resampling of the bootstrap. This is repeated 5000 times to generate a distribution of parameter values. From this distribution a standard deviation of each estimated parameter can be calculated, and a 95% confidence interval as the 0.025 and 0.975 quantiles of the distribution.

Outliers

On preliminary exploratory analysis, and as part of our quality assurance effort, we identified extreme outliers in the database as any value > Q3 + 4\*IQR, where Q3 is the third quartile for the empirical distribution and IQR is the interquartile range (Q3-Q1). First we determined if the outlier was due to an error in rounding, units or data entry. For those values that were not entered in error, we estimated distributions both with and without the value of the identified outlier. Below we give results for distributions estimated without outliers. All distribution estimates, both with and without outliers, are presented in the Supplementary Material.

Sample distributions

For the purposes of demonstrating comparisons of distributions among fuel types and EVT groups we present here distributions for EVT groups that represent eastern mixed hardwood forests (682 yellow birch-sugar maple, 655 beech-maple-basswood and 666 eastern floodplain) and conifer forests (683 peatland, 631 ponderosa pine, and 625 Douglas-fir, ponderosa pine and lodgepole pine). For these EVT groups we present distributions for total tree loading, coarse woody debris loading (downed wood > 7.6 cm diameter), duff loading, and litter loading. These were chosen because they represent major EVT groups across the continental US and because they also had sufficient representation of the chosen fuel types to estimate distributions.

To place the fuel loading distributions in the context of smoke management and emissions estimation, we made use of existing mapped surface fuel loading values (CWD, litter and duff) within LANDFIRE (2008 refresh – add citation) and compared them to fitted distributions for three common existing vegetation type groups in the eastern US and three in the western US and Alaska. These EVT groups were selected based on having good to excellent fuel loading distribution fits for coarse wood (CWD), litter and duff. Because fuel loading estimation is of particular importance for emissions calculations, we used Consume version 5.0 algorithms (Prichard et al. in prep) to calculate consumption and PM2.5 emissions. Mapped values were obtained for sample study areas for each EVT group (Table S2) by selecting an area that was dominated by a EVT Group (e.g., Harvard Forest, MA for 682 Yellow birch-sugar maple and Loomis State Forest, WA for 625 Douglas-fir – ponderosa pine – lodgepole pine forest) and selecting the most common FCCS fuelbed and fuel loading model that was used to represent that EVT group. Emissions were modeled in Consume using dry fuel moistures (15% for CWD, 11% for litter and 30% for duff). We then compared the point FCCS loading and associated emissions values to a 95% (or 75%?) prediction interval using the database distribution estimate.

**Results**

Database coverage

Of the 198 Landfire (2014) EVT groups, the database contains records for 134 EVT groups. Of those, 68 EVT groups had sufficient entries to estimate at least one fitted distribution of a fuel category (Table S1). Based on broad physiognomic or land use category, the majority of land area in the United States is forest and woodland (32%) followed by shrublands (19%), agriculture (17%) and non-vegetated pixels (16%). However, the percentage of EVT Groups with sufficient record counts for distribution fitting is highly skewed toward forest and woodland EVT groups (70%) with only 22% and 13% of shrubland and grassland EVT groups represented by at least one fitted distribution. Of the forest and woodland EVT Groups there was higher representation of fitted distributions for coniferous forests (78%) than broadleaf forests (68%) and mixed forests (63%).

**Table 4**: Percentage of total US land area by physiognomic class or agriculture and the percentage representation of each class with at least one fitted fuel loading distribution.

|  |  |  |
| --- | --- | --- |
|  | **% Area** | **% Representation** |
| Agriculture | 17% | 0% |
| Barren/developed/water | 16% | 0% |
| Forest and woodland | 32% | 70% |
| Grassland and tundra | 13% | 13% |
| Shrubland | 19% | 22% |
| Wetland and marsh | 2% | 13% |

Empirical distribution estimates

The results of all distributions that were estimated are given in supplementary material. Here we give example results for 6 representative EVT groups and fuel types. For eastern hardwood forests common in the continental US, tree loading is best represented by a gamma distribution (with nearly no zeroes). Observed values range from near 0 to near 500 Mg ha-1, with variability depending on EVT group. The example fits given here are all classified as at least “good” or “excellent.”

A common characteristic of many distributions is the high proportion of zero values and high coefficients of variation (CV). The presence and amounts of CWD and duff are particularly variable across records and leads to high CV in distribution estimates. These two categories also have a broad range in reported biomass values. For example, in the 3 sample mixed hardwood forest distributions (Figure 1), CWD ranges from 0-50 Mg/ha (Table 5), and the proportion of zero values is 5.6% in yellow birch and sugar maple (YB-SM) distributions, 14% in beech-maple-basswood (B-M-B) forests and nearly half of all records for eastern floodplain forests (47%). Mean CWD loadings are quite similar across all mixed hardwood sites as are the shapes of the distributions. In the three conifer forest distributions highlighted here (Figure 2), CWD ranges from 0 to nearly 150 Mg/ha, and the proportion of records with a zero value ranges from 8 to 21%. Mean CWD is greatest in Douglas-fir/ponderosa pine/lodgepole pine (DF-PP-LP) forests and lowest in peat forests. In each case, standard deviations meet or exceed estimated means (Table 5).

Relative to CWD, duff records have a relatively low proportion of zero values (4.4 to 7.3%). Duff loadings range widely from 0 to 250 Mg/ha in peatland forests and 0 to >100 Mg/ha in the other sample conifer forests. Mean duff loading is markedly higher in peatland forests (73 Mg/ha) compared to the other conifer forests (16-20 Mg/ha), and as with CWD, standard deviations are high. Presence of duff is more variable in the 3 sample mixed hardwood forests than conifer forests with the proportion of records without duff ranging from 4.5 to 23%. Mean duff loading ranges from 8 to 23 Mg/ha in mixed hardwood EVT groups, and standard deviations either exceed or are close to the estimated mean.

Litter and total aboveground biomass (TAB) of trees have less than 5% of records with zero values in the database records. Tree TAB ranges from 0-500 Mg/ha for conifer forests with highest mean tree biomass in DF-PP-LP forests (134 Mg/ha) compared to 75 Mg/ha for the other forest types. Mean tree TAB is comparable for mixed hardwood sites, ranging from 91 Mg/ha in eastern floodplain forests to 135 Mg/ha in YB-SM forests and high standard deviation and CV values.

Comparison with mapped estimates

Mapped fuel loading values from FCCS and FLMs across the 6 representative forest types often fall within fitted distributions but with notable exceptions. For CWD, FCCS fuelbed values match mean distribution fits for eastern hardwoods but are high for peatland forests, exceeding the Q3 value, and are also high for ponderosa pine forests (Figure 4). In contrast, the FCCS CWD values for DF-PP-LP forests falls below the Q1 value for that EVT group. CWD values for mapped FLMs are low across all EVT Groups, either at or below the Q1 value. When translated into PM2.5 emissions under a dry fuel moisture scenario, mapped FCCS values produce similar emissions estimates to the mean (Q2) distribution fit for the eastern hardwoods but are high for peatland and ponderosa pine forests and low for DF-PP-LP forests. Using the FLM CWD values underpredicts emissions relative to fitted distributions. For example, for the DF-PP-LP EVT group, using the mapped FCCS value to estimate emissions would predict 292 kg/ha vs 78 kg/ha from the FLM mapped value and 696 for the mean fitted distribution (Q2) value.

Mapped FCCS and FLM estimates of litter generally are much lower than Q1 values for the 6 representative forest types. However the FLM estimate for peatland litter is very high and nearly matches the Q3 value. Because mapped litter estimates are generally lower than even the Q1 values of fitted distributions, they are likely to underestimate PM25 emissions.

Eastern hardwood forests have low amounts of duff, reflected in fitted distribution values as well as mapped FCCS and FLM values. Mapped FCCS and FLM loading values for ponderosa pine and DF-PP-LP forests are both well within the range of distributions. However, the FCCS value for peatland forests is nearly four times that of the Q3 loading value. The FLM value also exceeds the Q3 value but is still within the fitted distribution. Because duff emissions factors are much higher than for CWD and litter, the consequences of errors in loading estimates are also greater. This is revealed for the peatland forests with even greater disparities in PM25 emissions than in loading values.

**Table 5**: Empirical distribution summaries for example fuel types and EVT groups. Prop 0 gives the proportion of entries with a value of zero loading for that fuel and EVT group. n gives the number of entries > 0 in the database for that fuel and EVT group. Quartiles (Q1, Q2, Q3; 25th, 50th, and 75th percentiles, respectively) are given in Mg ha-1 and represent those quantities for all entries > 0 for that fuel and EVT group. CV is the coefficient of variation, calculated as the standard deviation divided by the mean value for all entries > 0 for that fuel and EVT group. TAB = total aboveground biomass, CWD = coarse woody debris.

| **EVT**  **ID** | **EVT Group Name** | **Fuel type** | **Prop 0** | **n** | **Q1** | **Q2** | **Q3** | **CV** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 655 | Beech-maple-basswood forest | | | | | | | |
|  |  | Tree TAB | 0 | 159 | 76.1 | 121.69 | 169.55 | 0.51 |
|  |  | CWD | 0.142 | 139 | 2.45 | 6.34 | 12.89 | 0.98 |
|  |  | Duff | 0.148 | 138 | 2.3 | 5.69 | 17.01 | 1.21 |
|  |  | Litter | 0.018 | 161 | 6.47 | 9.32 | 13.28 | 0.52 |
| 682 | Yellow birch-sugar maple forest | | | | | | |  |
|  |  | Tree TAB | 0 | 387 | 87.23 | 127.16 | 176 | 0.52 |
|  |  | CWD | 0.056 | 370 | 3.53 | 7.42 | 13.44 | 1.05 |
|  |  | Duff | 0.045 | 378 | 6.22 | 17.93 | 33.69 | 0.9 |
|  |  | Litter | 0.005 | 400 | 5.85 | 8.85 | 12.7 | 0.53 |
| 666 | Eastern floodplain forest | | | | | | |  |
|  |  | Tree TAB | 0.003 | 1513 | 32.34 | 71.3 | 129.44 | 0.86 |
|  |  | CWD | 0.474 | 813 | 1.89 | 4.69 | 11.66 | 1.63 |
|  |  | Duff | 0.23 | 1164 | 2.3 | 3.45 | 7.75 | 1.7 |
|  |  | Litter | 0.039 | 1453 | 3.15 | 6.95 | 11.7 | 0.82 |
| 683 | Peatland forest | | | | | | | |
|  |  | Tree TAB | 0 | 68 | 40.62 | 75.15 | 104.48 | 0.59 |
|  |  | CWD | 0.162 | 57 | 1.57 | 4.26 | 7.84 | 1.09 |
|  |  | Duff | 0.049 | 58 | 25.56 | 58.72 | 86.26 | 0.87 |
|  |  | Litter | 0.016 | 60 | 3.55 | 7.11 | 11.58 | 0.89 |
| 631 | Ponderosa pine forest and woodland | | | | | | |  |
|  |  | Tree TAB | 0.013 | 301 | 31.59 | 66.85 | 105.4 | 0.75 |
|  |  | CWD | 0.211 | 946 | 3.58 | 9.95 | 19.38 | 1.38 |
|  |  | Duff | 0.044 | 475 | 4.55 | 10.65 | 21.45 | 1.07 |
|  |  | Litter | 0.003 | 578 | 3.67 | 6.34 | 11.83 | 1.41 |
| 625 | Douglas-fir, ponderosa pine and lodgepole pine forest | | | | | | |  |
|  |  | Tree TAB | 0.043 | 90 | 61.84 | 108.34 | 190.28 | 0.85 |
|  |  | CWD | 0.081 | 763 | 7.54 | 17.2 | 36.64 | 1.6 |
|  |  | Duff | 0.073 | 178 | 6.56 | 14.59 | 30.05 | 0.87 |
|  |  | Litter | 0.013 | 153 | 4.67 | 7.32 | 12.57 | 0.92 |

**DISCUSSION**

Fuels maps are used as data inputs to numerous modelling applications, including consumption and emissions models such as the First Order Fire Effects Model (Reinhardt et al. 1998) and Consume (Prichard et al. in press), wildland fire behavior prediction tools such as FLAMMAP (Finney 2006) and smoke dispersion modeling tools and frameworks such as the Wildland Fire Emissions Information System (French et al. 2011) and BlueSky (Larkin et al. 2014). For many modeling studies of biomass and climate, the importance of incorporating uncertainty is the foundation of simulations. For example, coarse-scale dynamic vegetation models draw inputs from probability distributions in order to model stochastic processes of fire and climate (Quillet et al. 2010, Shankar et al. 2018). Bootstrapping is also used to understand the effect of sampling variability on model predictions; for example, Gregg and Hummel (2002) used bootstrapped tree lists to evaluate impacts on Forest Vegetation Simulator projections (Gregg and Hummel 2002). However, to date, distributions of mapped fuels have not been available for simulating fire behavior, effects and smoke production. Despite the acknowledged variability of fuels at multiple spatial scales (Keane et al. 2012), there are currently no products that incorporate uncertainty in estimating the biomass of wildland fuels in North America.

The fuel loading database offers a compilation of existing data sources on surface and canopy fuels for the United States and Canada. As expected, we observed a wide range in fuel loadings by major fuel category (aboveground tree biomass, shrubs, herbs, CWD, litter and duff), and distribution fits were generally best fitted with either a gamma or lognormal distribution, right-skewed distributions representing a majority of low values with long right tails to higher values. Of the 134 EVT Groups supported in the database, we are able to present at least one fitted distribution for 64 EVT groups. Fitted distributions can be used to provide random draws from the distribution fits or values that range across percentiles, including the 25th, 50th and 75th percentile values presented in this paper.

Database development revealed substantial bias in forest records relative to non-forest types. Given that forests have been intensively measured for timber resources and other forest management goals, this is not particularly surprising. However, because of the potential importance of these datasets for informing uncertainty in carbon mapping and emissions inventories, the data gaps revealed in this study may justify fuel characterization in non-forest vegetation types and inform future sampling campaigns. Some of the grassland and shrubland EVT groups may have low record counts because vegetation descriptions were insufficient for specific group assignments. For example, of the 5 most common shrubland types, we only had sufficient record counts to provide distribution fits for big sagebrush and mesquite shrublands; salt desert, willow and developed upland shrubland EVT groups are all sparsely represented in the fuel loading database. Agricultural EVT groups comprise a large percentage of US land area and are not currently represented in the database. Fuel loading records are needed to represent pasturelands, row crows and wheat fields in particular. Of the major forest types, plantations and white spruce forests and woodlands lack sufficient records. Another issue with low representation of many EVT groups is the lack of specification in vegetation types and associated descriptive data. Of the 6 most common grassland types (developed, shortgrass prairie, mixed grass prairie, upland herbaceous and tussock tundra and grassland), only mixed grass prairie and grassland EVT groups are currently supported in the database. There are over 500 generic grassland records, and a major improvement to the database would be to improve the specificity of these records so that they could be binned within specific grassland EVT groups. In some cases, this might be possible by additional geospatial analysis of plot locations and assignment of likely vegetation.

Database uses

Even with the data gaps identified in this manuscript, the fuel loading dataset should be immediately useful for applications including carbon, fire hazard assessments, and emissions inventories. The database and fitted distributions allow for three major advances in fuel characterization:

1) By using distributions of fuel loading for a vegetation type rather than a point mapped estimate, a credible interval of emissions estimates or carbon accounting can be generated. The ability to calculate uncertainty bounds on model predictions gives users a better idea of a plausible range of model outputs rather than a single point estimate.

2) The distributions of fuel loading for major vegetation types can also be used to evaluate potential errors in point estimates given in current map products. Mapped values can be assessed by comparing them to distribution estimates (Figure) to determine if the mapped values are representative of known EVT distributions. For example, the FCCS mapped duff value for peatland forests is clearly not in the center of the distribution estimated here (Figure). Given this discrepancy, we would recommend that the representative FCCS fuelbed (FB279: Black spruce-larch-northern white cedar forest) be modified to a more reasonable estimate such as the reported Q2 value for peatland forests. Additional comparisons may point to other improvements in mapped products.

3) Providing distribution fits by major fuel category can also help inform sensitivity and uncertainty analysis of fuels as inputs for evaluating specific management objectives. For example, we can use the distribution estimates to understand how uncertainty in fuel loadings propagate to uncertainty in wildfire emissions. For a given EVT group, a sensitivity analysis of fuel loading inputs to emissions models can identify the fuel types that contribute most to uncertainty in model predictions (e.g., coarse wood or duff). By identifying the most sensitive fuel loading types, improved quantification can be prioritized for these fuel types.

Conclusions

This studyFor model applications including carbon inventories, fire hazard analysis and emissions estimation, it is insufficient to provide point estimates of fuel loadings. Individual point estimates are unlikely to be absolutely correct but rather represent the best scalar value for a given pixel at a given time. As with all efforts in estimation, it is best to provide both a point estimate and a credible interval around that estimate. In this study, we found that even with the best -available datasets there is tremendous uncertainty in fuel loading values, which propagates to uncertainty in associated model applications. Model prediction the use point estimates of fuel loading are therefore vastly under-representing the uncertainty associated with those model predictions. Even with the large number of records in our database, there are also substantial gaps in measured fuel loadings by vegetation type and region with notable gaps in many non-forest EVT groups. When compiling the database we considered grouping the observations by Existing Vegetation Type, but quickly realized there was insufficient coverage of observations to estimate distributions by vegetation type. This led us to aggregate by Existing Vegetation Group, and we see substantial variability in fuel loading when data are grouped at this level. Increased coverage of observations by vegetation type would likely reduce the variability in loading and associated uncertainty.

The fuel loading database is a work in progress and will benefit from additional records, particularly for under-represented fuel types. As such, it offers guidance on where additional field inventories and fuel characterization would be particularly beneficial. The existing fuel loading distributions for 65 EVT Groups will be useful to assessing how representative existing mapped estimates of fuel loadings are and also for calculating uncertainty in fuel loading estimates and highly correlated estimates of potential wildland fire emissions and carbon stores.

**References**

Anderson, H.E., 1982. Aids to determining fuel models for estimating fire behavior. INT-GTR-122. US Forest Service Intermountain Forest and Range Experiment Station, Ogden, UT.

Blackard, J.A., Finco, M.V., Helmer, E.H. et al. 2007. Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sensing of Environment 112:1658-1677.

Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L. and Kimes, D. 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment, 112:3876–3890.

Congalton, R.G., Gu, J., Yadav, K., Thenkabail, P. and Ozdogan, M. 2014. Global land cover mapping: a review and uncertainty analysis. Remote Sensing 6:12070-12093.

Delignette-Muller, M.L. and Dutang, C. 2015. fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64:1-34. <http://www.jstatsoft.org/v64/i04>

Finney, M. A. 2006. An overview of FlamMap fire modeling capabilities. In: Fuels management—how to measure success: conference proceedings. 2006 March 28-30; Portland, Oregon. Proceedings RMRS-P-41. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 213-220.

French, N.H.F., W.J. de Groot, L.K. Jenkins, B.M. Rogers, E.C. Alvarado, B. Amiro, B. de Jong, S. Goetz, E. Hoy, E. Hyer, R. Keane, D. McKenzie, S.G. McNulty, B.E Law, R. Ottmar, D.R. Perez-Salicrup, J. Randerson, K.M. Robertson, and M. Turetsky. 2011. Model comparisons for estimating carbon emissions from North American wildland fire. Journal of Geophysical Research 116.

Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J. and Guo, Q. 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sensing 8:565.

Keane, R.E., Burgan, R. and van Wagtendonk, J. 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. International Journal of Wildland Fire 10:301-319.

Keane, R.E. 2013. Describing wildland surface fuel loading for fire management: a review of approaches, methods and systems. International Journal of Wildland Fire 22:51-62.

Keane, R.E., Herynk, J.M., Toney, C., Urbanski, S.P., Lutes, D.C. and Ottmar, R.D. 2013. Evaluating the performance and mapping of three fuel classification system using Forest Inventory and Analysis surface fuel measurements. Forest Ecology and Management 305:248-263.

Krasnow, K., Schoennagel, T. and Veblen, T.T. 2009. Forest fuel mapping and evaluation of LANDFIRE fuel maps in Boulder County, Colorado, USA. Forest Ecology and Management 257:1603-1612.

Lachenbruch, P.A. 2002. Analysis of data with excess zeros. Statistical Methods in Medical Research. 11, 297–302. doi:10.1191/0962280202sm289ra

Larkin, N.K., Raffuse, S.M. and Strand, T.M. 2014. Wildland fire emissions, carbon, and climate: US emissions inventories. Forest Ecology and Management 317:61-69.

Lutes, D.C., Keane, R.E. and Caratti, J. 2009. A surface fuel classification for estimating fire effects. International Journal of Wildland Fire 18:802-814.

Ottmar, R.D., Sandberg, D.V., Riccardi, C.L. and Prichard, S.J. 2007. An overview of the Fuel Characteristic Classification System – quantifying, classifying, and creating fuelbeds for resource planning. Canadian Journal of Forest Research 37:2383-2393.

Ottmar, R.D. 2014. Wildland fire emissions, carbon, and climate: modeling fuel consumption. Forest Ecology and Management 317:41-50.

Pierce, K.B. Jr., Ohmann, J.L., Wimberly, M.C., Gregory, M.J. and Fried, J.S. 2009. Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods. Canadian Journal of Forest Research 39:1901-1916.

Peterson, J., Lahm, P., Fitch, M., George, M., Haddow, D., Melvin, M., Hyde, J. and Eberhardt, E. (editors) 2018. NWCG smoke management guide for prescribed fire. National Wildfire Coordinating Group PMS 420-2, NFES 001279. CHECK REF FORMAT

Reinhardt, E.D., Keane, R.E., and Brown, J.K. 1997. First order ﬁre effects model: FOFEM. GTR-INT-344.US Forest Service, Rocky Mountain Research Station, Ogden, Utah,

Riley, K.L., Grenfell, I.C. and Finney, M.A. 2016. Mapping forest vegetation for the western United States using modified random forests imputation of FIA forest plots. Ecosphere 7:e01472.

Rollins, M.G. 2009. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. International Journal of Wildland Fire 18:235-249.

Rollins, M.G., Keane, R.E., and Parsons, R.A. 2004. Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. Ecological Applications 14:75-95.

Scott, J.H., Thompson, M.P., Calkin, D.E. 2013. A wildfire risk assessment framework for land and resource management. RMRS-GTR-315. US Forest Service Rocky Mountain Research Station. 83p.

Thurner, M.; Beer, C.; Santoro, M.; Carvalhais, N.; Wutzler, T.; Schepaschenko, D.; Shvidenko, A.; Kompter, E.; Ahrens, B.; Levick, S.R.; et al. Carbon stock and density of northern boreal and temperate forests. Global Ecology and Biogeography23:297–310.

Urbanski, S. 2014. Wildland fire emissions, carbon and climate: emissions factors. Forest Ecology and Management 317:51-60.

US EPA 2017 Profile of version 1 of the 2014 National Emissions Inventory.

<https://www.epa.gov/sites/production/files/2017-04/documents/2014neiv1_profile_final_april182017.pdf> [accessed June 21, 2018]

Beck, M. B. 1987. Water quality modeling: A review of the analysis of uncertainty. Water Resources Research 23:1393–1442.

Delignette-Muller, M., and C. Dutang. 2015. fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 64:1–34.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse. 1983. A Comparison of Alternative Models for the Demand for Medical Care. Journal of Business & Economic Statistics 1:115–126.

Gregg, T. F., and S. Hummel. 2002. Assessing Sampling Uncertainty in FVS Projections Using a Bootstrap Resampling Method.

Hanna, S. R. 1988. Air quality model evaluation and uncertainty. JAPCA 38:406–412.

Hummel, S., and P. Cunningham. 2006. Estimating variation in a landscape simulation of forest structure. Forest Ecology and Management 228:135–144.

Lachenbruch, P. A. 2002. Analysis of data with excess zeros. Statistical Methods in Medical Research 11:297–302.

Lecomte, J. B., H. P. Benoît, S. Ancelet, M. P. Etienne, L. Bel, and E. Parent. 2013. Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. Methods in Ecology and Evolution 4:1159–1166.

Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. American Statistical Journal June:399–402.

Massart, P. 1990. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. The Annals of Probability 18:1269–1283.

O’neill, R. V., and R. H. Gardner. 1979. Sources of Uncertainty in Ecological Models. Pages 447–463 in B. P. Zeigler, M. S. Elzas, G. J. Klir, and T. I. Oren, editors. Methodology in Systems Modelling and Simulation. North-Holland Publishing Company, Amsterdam. New York. Oxford, New York, US.

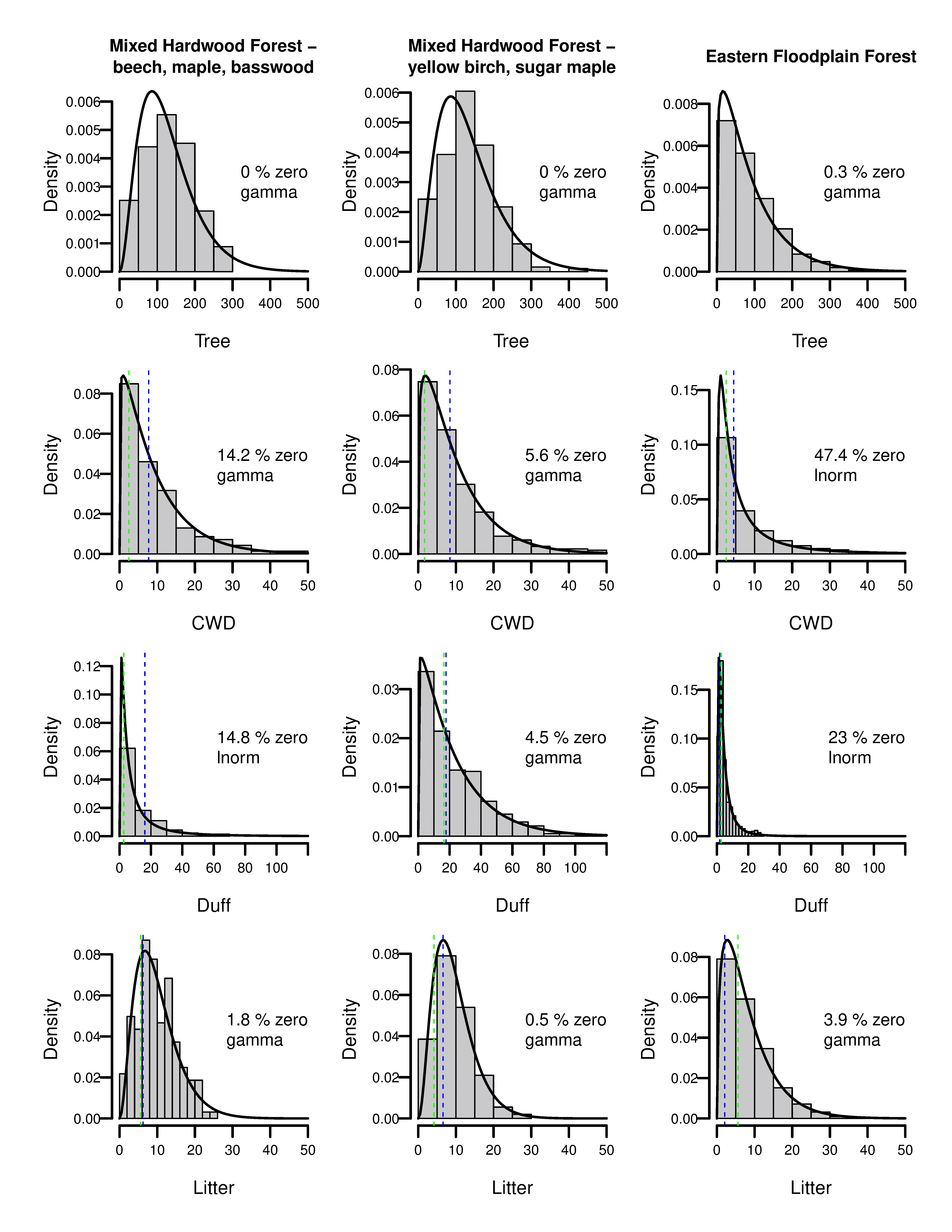
Prichard, S. J., E. C. Karau, R. D. Ottmar, M. C. Kennedy, J. B. Cronan, C. S. Wright, and R. E. Keane. 2014. Evaluation of the CONSUME and FOFEM fuel consumption models in pine and mixed hardwood forests of the eastern United States. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 44:784–795.

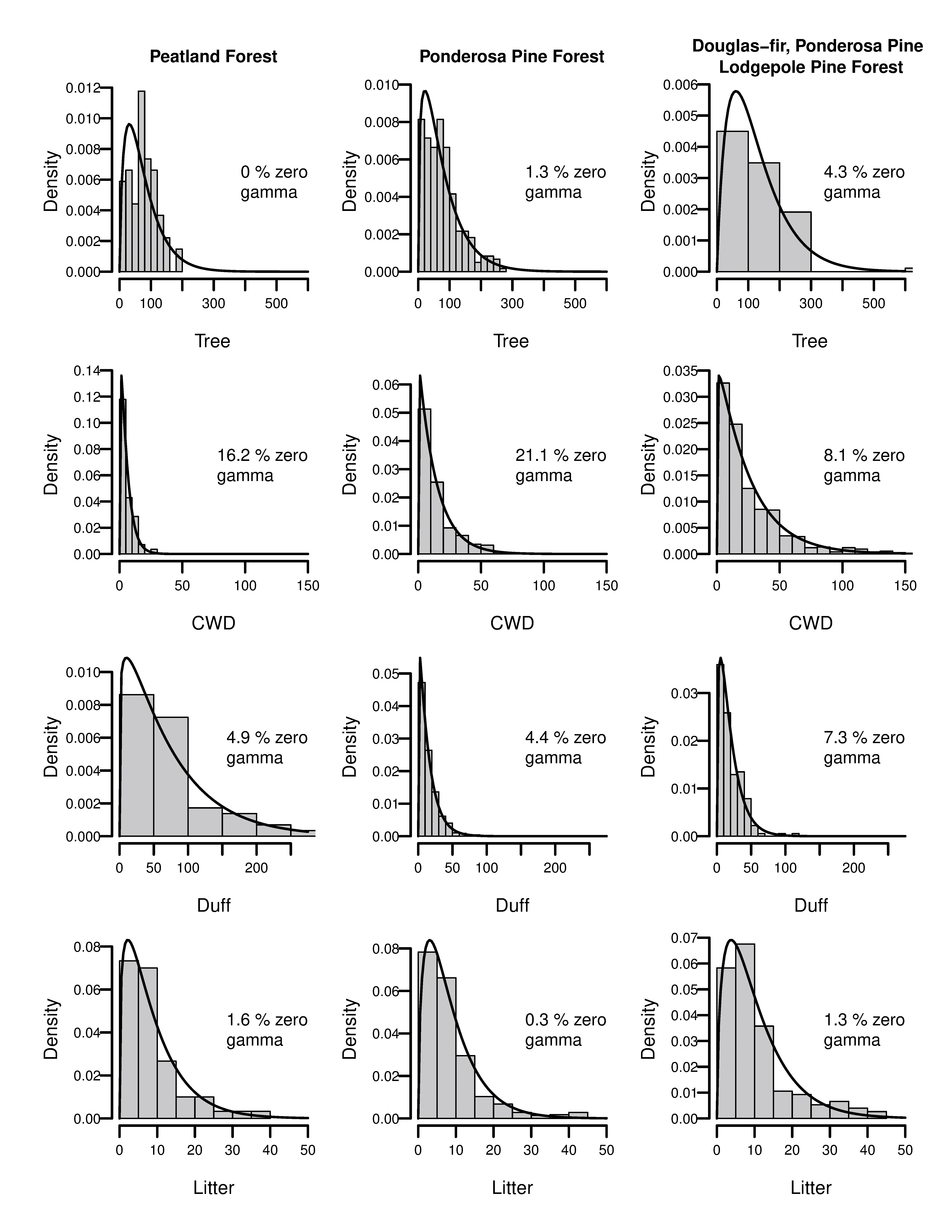
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, A. P., and R. E. Froese. 2004. Model validation using equivalence tests. Ecological Modelling 176:349–358.

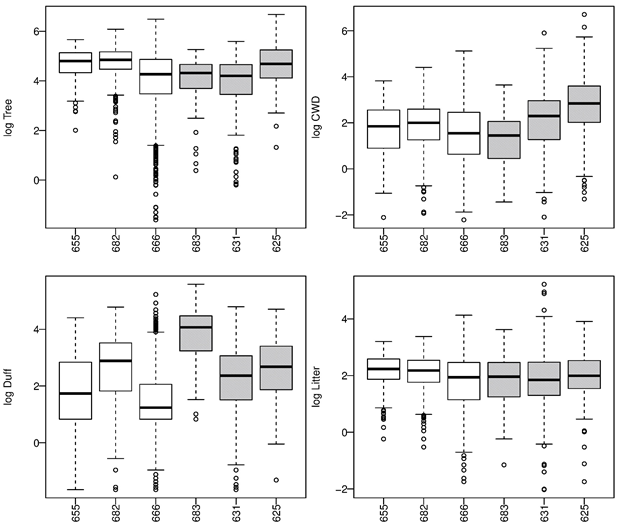
Turley, M. C., and E. D. Ford. 2009. Definition and calculation of uncertainty in ecological process models. Ecological Modelling 220:1968–1983.

Welsh, A. H., R. B. Cunningham, C. F. Donnelly, and D. B. Lindenmayer. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88:297–308.

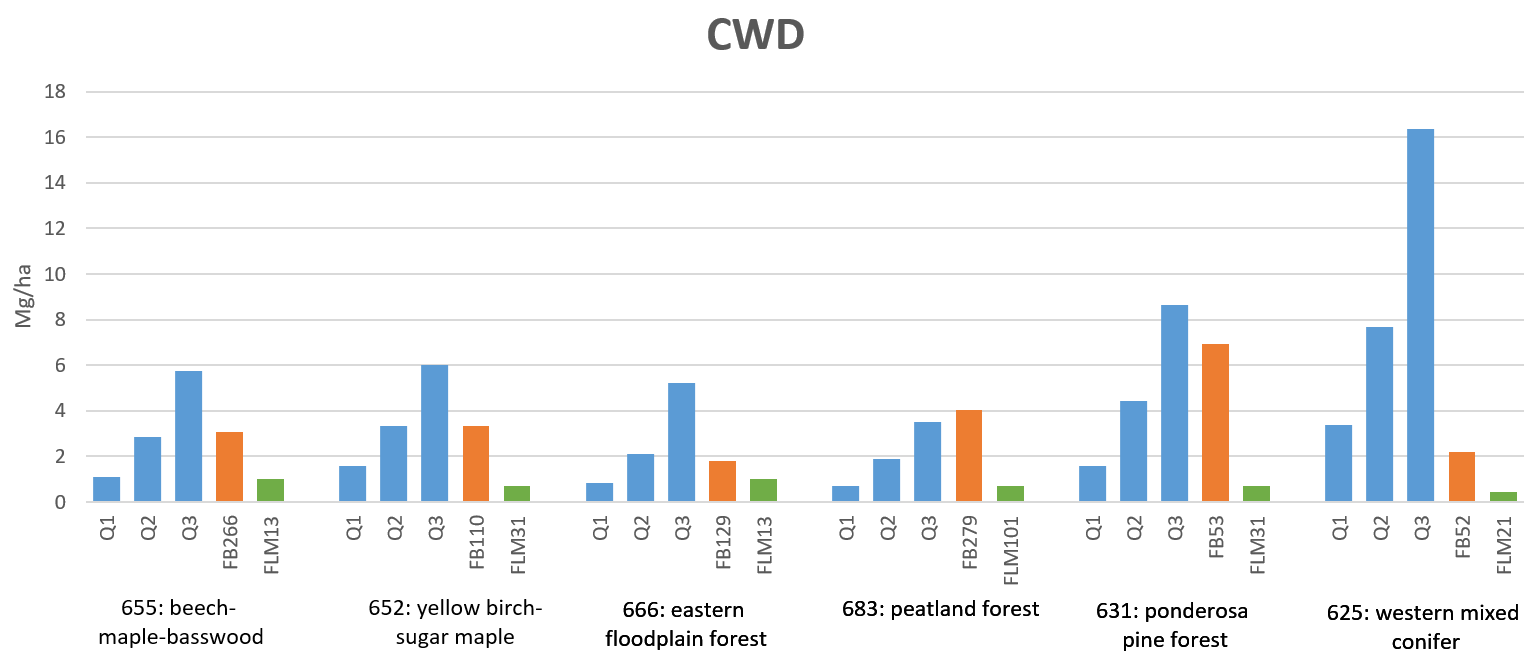
**Figure 1**: Fitted distributions for aboveground tree biomass (Tree), coarse woody debris (CWD) and organic soil layers (Duff and Litter) of three sample eastern mixed hardwood existing vegetation type groups. Green vertical lines give mapped values for FLM, blue for FCCS.

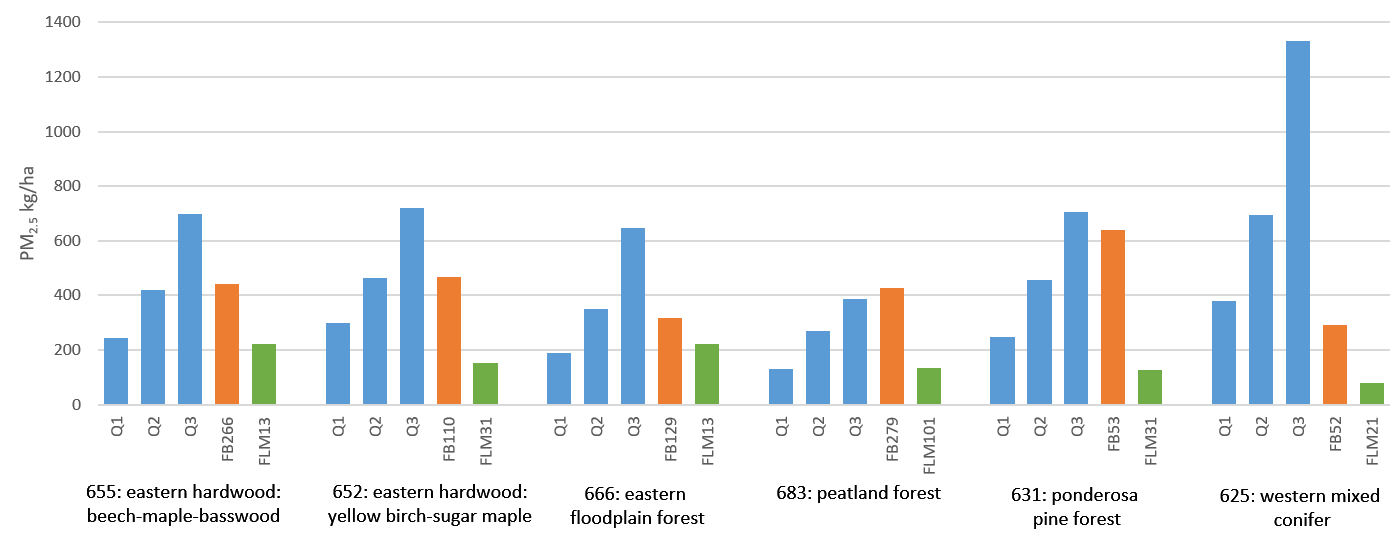


**Figure 2**: Fitted distributions for aboveground tree biomass (Tree), coarse woody debris (CWD) and organic soil layers (Duff and Litter) of three sample western conifer existing vegetation type groups.

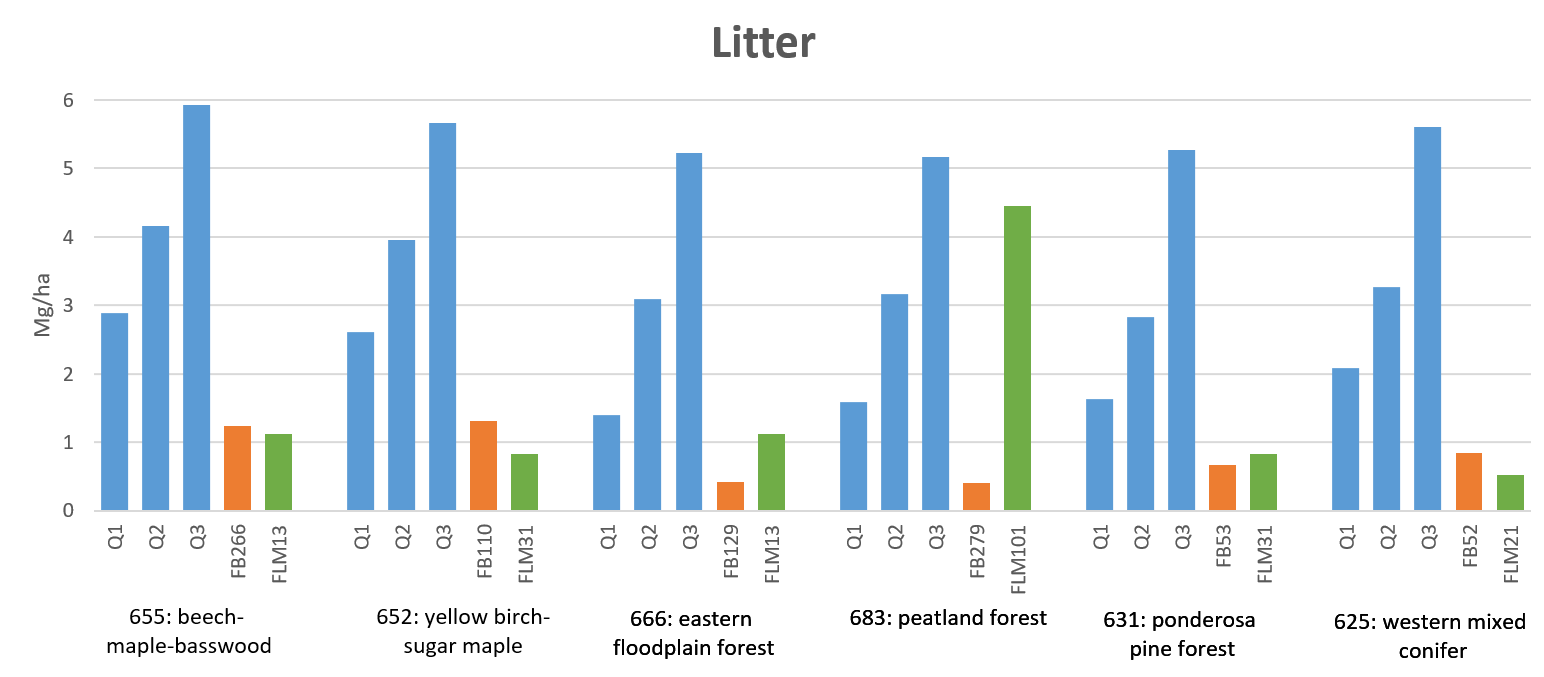


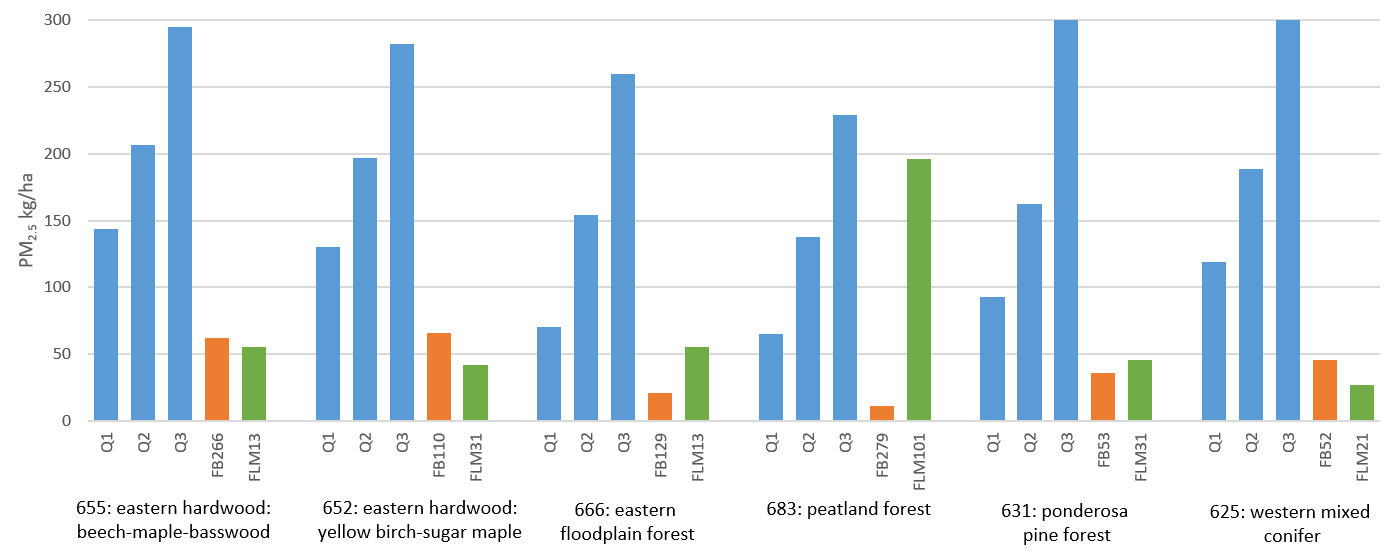
**Figure 3**: Boxplot comparison of example EVT groups and fuel loading. All values are log transformations of Mg ha-1, to aid comparisons. White boxes are hardwood, grey boxes conifer. Takeaways: for some fuels distributions are very similar (e.g., trees, litter), and for others there are some differences in distributions (CWD, duff). There is a lot of variability regardless of EVT group.



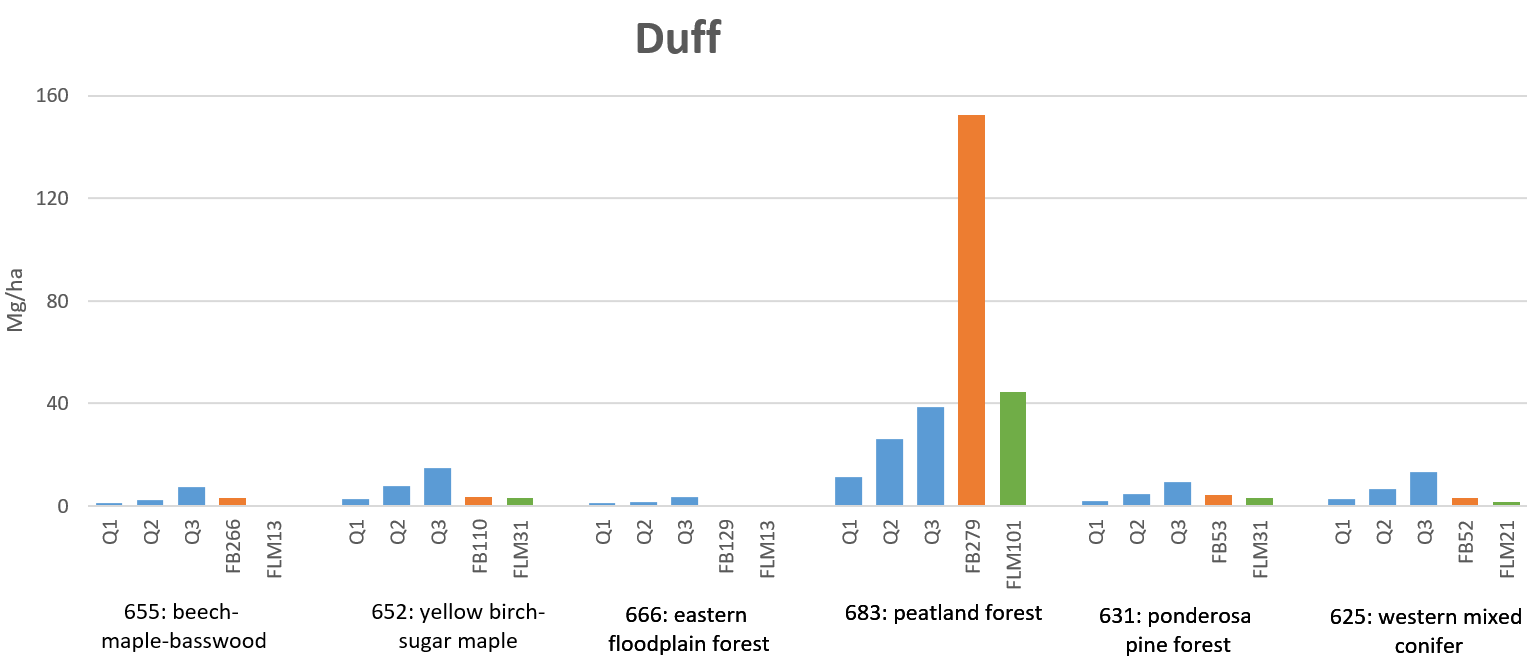


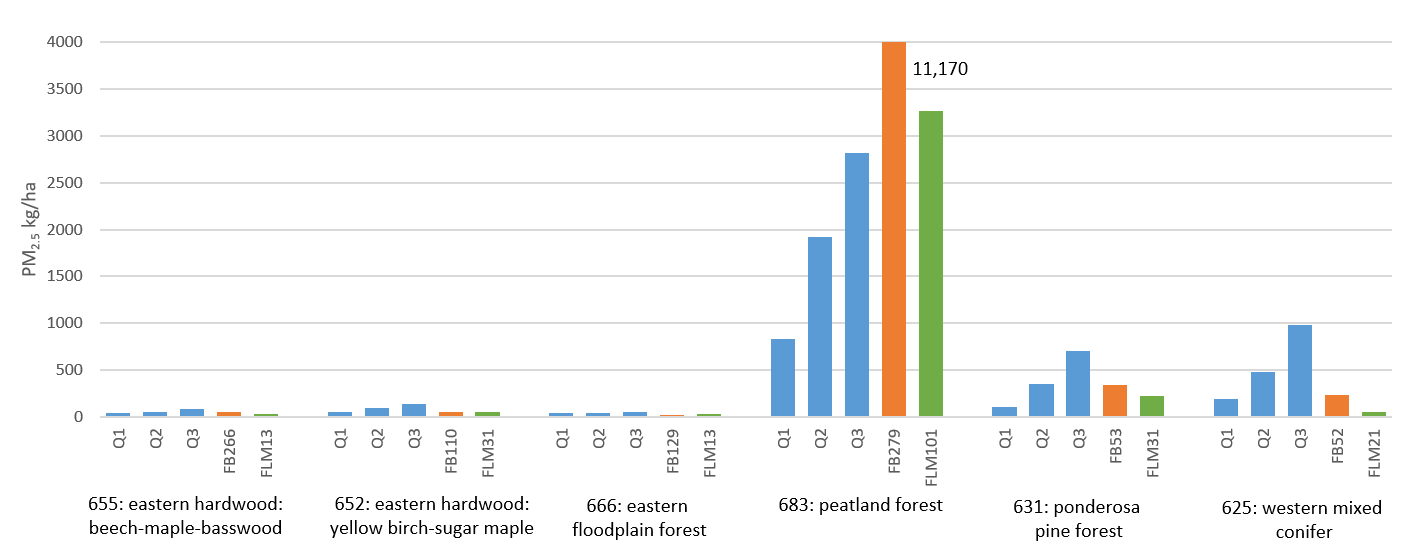
**Figure 4:** Comparison of 25th (Q1), 50th (Q2) and 75th (Q3) percentile distributions of coarse wood (CWD) loading values (A) and calculated PM2.5 emissions (B) with mapped values for that EVT group, obtained from sample areas within the LANDFIRE 2008 Fuel Characteristics Classification System and Fuel Loading Model layers.





**Figure 5:** Comparison of 25th (Q1), 50th (Q2) and 75th (Q3) percentile distributions of litter loading values (A) and calculated PM2.5 emissions (B) with mapped values for that EVT group, obtained from sample areas within the LANDFIRE 2008 Fuel Characteristics Classification System and Fuel Loading Model layers.





**Figure 6:** Comparison of 25th (Q1), 50th (Q2) and 75th (Q3) percentile distributions of duff loading values (A) and calculated PM2.5 emissions (B) with mapped values for that EVT group, obtained from sample areas within the LANDFIRE 2008 Fuel Characteristics Classification System and Fuel Loading Model layers.

**Table S1**: Summary of EVT groups with sufficient records to support at least one fitted distribution. Percentage of land area is calculated as percentage of pixels represented by each EVT Group, based on the total pixels in the LANDFIRE EVT vegetation layer.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **EVT Group ID** | **EVT Group Name** | **Dominant species** | **% Land Area** | **NumStudyPoints** |
| 604 | Big Sagebrush Shrubland and Steppe | *Artemesia tridentata* | 4.428% | 144 |
| 646 | Mixedgrass Prairie |  | 3.140% | 46 |
| 613 | Desert Scrub |  | 2.048% | 53 |
| 630 | Pinyon-Juniper Woodland | *Pinus edulis*  *Juniperus* spp. | 1.736% | 626 |
| 682 | Yellow Birch-Sugar Maple Forest | *Betula alleghaniensis*  *Acer saccharum* | 1.639% | 415 |
| 660 | White Oak-Red Oak-Hickory Forest and Woodland | *Quercus alba*  *Q. rubra, Carya* spp. | 1.633% | 1161 |
| 655 | Beech-Maple-Basswood Forest | *Fagus grandifolia*  *Acer spp.*  *Tilia americana* | 1.619% | 167 |
| 617 | Grassland |  | 1.343% | 116 |
| 624 | Mesquite Woodland and Scrub | *Prosopis* spp. | 1.255% | 1309 |
| 683 | Peatland Forests | *Picea glauca*  *P. mariana* | 1.152% | 69 |
| 666 | Eastern Floodplain Forests |  | 1.072% | 1548 |
| 740 | Ruderal Forest |  | 1.064% | 347 |
| 758 | Black Spruce Forest and Woodland | Picea mariana | 0.961% | 131 |
| 631 | Ponderosa Pine Forest, Woodland and Savanna | *Pinus ponderosa* | 0.938% | 1426 |
| 661 | Chestnut Oak Forest and Woodland | *Quercus prinus* | 0.933% | 276 |
| 694 | Atlantic Swamp Forests |  | 0.883% | 235 |
| 664 | Chestnut Oak-Virginia Pine Forest and Woodland | *Quercus prinus*  *Pinus virginiana* | 0.864% | 51 |
| 685 | Pine-Hemlock-Hardwood Forest | *Pinus spp.*  *Tsuga canadensis*  *Acer* spp. | 0.825% | 145 |
| 625 | Douglas-fir-Ponderosa Pine-Lodgepole Pine Forest and Woodland | *Pseudotsuga menziesii*  *Pinus ponderosa*  *P. contorta* | 0.797% | 893 |
| 635 | Western Riparian Woodland and Shrubland |  | 0.760% | 136 |
| 639 | Spruce-Fir Forest and Woodland | *Picea glauca*  *Abies balsamea* | 0.706% | 995 |
| 662 | Post Oak Woodland and Savanna | *Quercus stellata* | 0.703% | 355 |
| 614 | Douglas-fir Forest and Woodland | *Pseudotsuga menziesii* | 0.684% | 990 |
| 693 | Spruce-Fir-Hardwood Forest | *Picea glauca*  *P. mariana*  *Abies balsamea* | 0.651% | 304 |
| 643 | Douglas-fir-Grand Fir-White Fir Forest and Woodland | *Pseudotsuga menziesii*  *Abies concolor*  *A. grandis* | 0.643% | 1056 |
| 756 | Birch-Aspen Forest | *Betula* spp.  Populus tremuloides | 0.569% | 48 |
| 615 | Douglas-fir-Western Hemlock Forest and Woodland | *Pseudotsuga menziesii*  *Tsuga heterophylla* | 0.507% | 1445 |
| 696 | Juniper-Oak | *Juniperus* spp.  *Quercus* spp. | 0.502% | 140 |
| 677 | Longleaf Pine Woodland | *Pinus palustris* | 0.458% | 114 |
| 668 | Eastern Small Stream Riparian Forests |  | 0.414% | 123 |
| 680 | Sweetgum-Water Oak Forest | *Liquidambar styraciflua*  *Quercus nigra* | 0.406% | 457 |
| 612 | Deciduous Shrubland |  | 0.392% | 93 |
| 684 | Pine Flatwoods | *Pinus palustris*  *P. taeda* | 0.381% | 1132 |
| 649 | Tallgrass Prairie | *Andropogon geradii*  *Schizachyrium scoparium* | 0.375% | 88 |
| 602 | Aspen Forest, Woodland, and Parkland | *Populus tremuloides* | 0.340% | 138 |
| 622 | Lodgepole Pine Forest and Woodland | *Pinus contorta* | 0.312% | 882 |
| 610 | Conifer-Oak Forest and Woodland |  | 0.247% | 187 |
| 690 | Shortleaf Pine-Oak Forest and Woodland | *Pinus echinata*  *Quercus* spp. | 0.245% | 137 |
| 689 | Shortleaf Pine Woodland | *Pinus echinata* | 0.225% | 125 |
| 626 | California Mixed Evergreen Forest and Woodland | *Abies spp.*  *Calocedrus decurrens*  *Pinus spp.* | 0.209% | 462 |
| 603 | Aspen-Mixed Conifer Forest and Woodland | *Populus tremuloides*  *Pinus spp.*  *Abies spp.* | 0.207% | 105 |
| 640 | Subalpine Woodland and Parkland |  | 0.200% | 281 |
| 675 | Inland Marshes and Prairies |  | 0.199% | 49 |
| 627 | Mountain Hemlock Forest and Woodland | *Tsuga mertensiana* | 0.196% | 350 |
| 638 | Sitka Spruce Forest | *Picea sitchensis* | 0.191% | 132 |
| 667 | White Oak-Beech Forest and Woodland | *Quercus alba*  *Fagus grandifolia* | 0.190% | 147 |
| 642 | Western Hemlock-Silver Fir Forest | *Tsuga heterophylla*  *Abies amabilis* | 0.182% | 1118 |
| 652 | Aspen-Birch Forest | *Populus tremuloides*  *Betula spp.* | 0.170% | 346 |
| 663 | Black Oak Woodland and Savanna | *Quercus kelloggii* | 0.136% | 203 |
| 676 | Jack Pine Forest | *Pinus banksiana* | 0.117% | 89 |
| 645 | Western Red-cedar-Western Hemlock Forest | *Thuja plicata*  *Tsuga heterophylla* | 0.117% | 368 |
| 633 | Red Fir Forest and Woodland | *Abies magnifica* | 0.109% | 169 |
| 632 | Red Alder Forest and Woodland | *Alnus rubra* | 0.103% | 90 |
| 629 | Western Oak Woodland and Savanna | *Quercus spp.* | 0.101% | 162 |
| 620 | Juniper Woodland and Savanna | *Juniperus spp.* | 0.098% | 617 |
| 665 | Red Pine-White Pine Forest and Woodland | *PInus resinosa*  *P. strobus* | 0.096% | 229 |
| 628 | Mountain Mahogany Woodland and Shrubland | *Cercocarpus spp.* | 0.089% | 69 |
| 657 | Cypress | *Taxodium distichum* | 0.053% | 55 |
| 695 | Virginia Pine Forest | *Pinus virginiana* | 0.049% | 84 |
| 659 | Bur Oak Woodland and Savanna | *Quercus macrocarpa* | 0.034% | 71 |
| 656 | Texas Live Oak | *Quercus fusiformis* | 0.034% | 236 |
| 686 | Pitch Pine Woodlands | *Pinus rigida* | 0.033% | 81 |
| 697 | Loblolly Pine Forest and Woodland | *Pinus taeda* | 0.024% | 1320 |
| 707 | Introduced Upland Vegetation-Treed |  | 0.021% | 61 |
| 679 | Maritime Forest |  | 0.019% | 79 |
| 691 | Southern Scrub Oak | *Quercus spp.* | 0.004% | 48 |
| 672 | Hammocks |  | 0.003% | 33 |
| 644 | Western Larch Forest and Woodland | *Larix occidentalis* | 0.002% | 135 |

**Table S2**: Sampled LANDFIRE mapped fuels data to compare with EVT Group distributions. ADD ADDITIONAL METHODS HERE?

|  | **Fuel type** | **Q1** | **Q2** | **Q3** | **FCCS value** | **FLM** |
| --- | --- | --- | --- | --- | --- | --- |
| **655 Beech-maple-basswood forest** | | | | |  |  |
| Cuivre River State Park, MO | | | | | FB266: Sugar maple-basswood forest | 13: Moderate FWD, light to mod litter, light duff |
|  | Tree TAB | 76.1 | 121.69 | 169.55 | 105.1 | n/a |
|  | CWD | 2.45 | 6.34 | 12.89 | 6.9 | 2.23 |
|  | Litter | 6.47 | 9.32 | 13.28 | 2.79 | 2.5 |
|  | Duff | 2.3 | 5.69 | 17.01 | 7.2 | 1.2 |
| **682 Yellow birch-sugar maple forest** | | | | |  |  |
| Harvard Forest, MA | | | | | FB110: American beech-yellow birch-sugar maple forest | 31: Moderate litter, light duff, light logs |
|  | Tree TAB | 87.23 | 127.16 | 176 | 120.2 | n/a |
|  | CWD | 3.53 | 7.42 | 13.44 | 7.5 | 1.52 |
|  | Litter | 5.85 | 8.85 | 12.7 | 2.95 | 1.87 |
|  | Duff | 6.22 | 17.93 | 33.69 | 7.92 | 7.31 |
| **666 Eastern floodplain forest** | | | | |  |  |
| Confluence of the Mississippi and Illinois Rivers | | | | | FB129: Green ash-American elm forest | 13: Moderate FWD, light to mod litter, light duff |
|  | Tree TAB | 32.34 | 71.3 | 129.44 | 87.95 | n/a |
|  | CWD | 1.89 | 4.69 | 11.66 | 4 | 2.23 |
|  | Litter | 3.15 | 6.95 | 11.7 | 0.93 | 2.5 |
|  | Duff | 2.3 | 3.45 | 7.75 | 0.8 | 1.2 |
|  |  |  |  |  |  |  |
| **683 Peatland forest** | | | | | | |
| Hiawatha National Forest, MI | | | | | FB279: Black spruce-larch-northern white cedar forest | 101: Very heavy duff |
|  | Tree TAB | 40.62 | 75.15 | 104.48 | 45.05 | n/a |
|  | CWD | 1.57 | 4.26 | 7.84 | 9 | 1.61 |
|  | Litter | 3.55 | 7.11 | 11.58 | 0.91 | 9.99 |
|  | Duff | 25.56 | 58.72 | 86.26 | 342 | 99.99 |
|  |  |  |  |  |  |  |
| **631 Ponderosa pine forest and woodland** | | | | | | |
| Lake Roosevelt National Recreation Area | | | | | FB 53: Pacific ponderosa pine forest | 31: Moderate litter, light duff, light logs |
|  | Tree TAB | 31.59 | 66.85 | 105.4 | 56 | n/a |
|  | CWD | 3.58 | 9.95 | 19.38 | 15.5 | 1.52 |
|  | Litter | 3.67 | 6.34 | 11.83 | 1.5 | 1.87 |
|  | Duff | 4.55 | 10.65 | 21.45 | 10.2 | 7.31 |
| **625 Douglas-fir, ponderosa pine and lodgepole pine forest** | | |  |  | FB 52 | 21 |
| Loomis State Forest | | | | | Douglas-fir ponderosa pine forest | light logs, light duff |
|  | Tree TAB | 61.84 | 108.34 | 190.28 | 48.89 | n/a |
|  | CWD | 7.54 | 17.2 | 36.64 | 4.9 | 0.94 |
|  | Litter | 4.67 | 7.32 | 12.57 | 1.88 | 1.16 |
|  | Duff | 6.56 | 14.59 | 30.05 | 7.5 | 3.3 |