

JGR Biogeosciences

RESEARCH ARTICLE

10.1029/2019JG005083

Key Points:

- This study presents fitted distributions of wildland fuel biomass for major vegetation of the United States and Canada
- We found that even with the best available data sets there is tremendous uncertainty in biomass values, which propagates to uncertainty in model applications
- Uncertainty estimates in biomass can be used for carbon inventories, fire hazard analysis, and emission modeling

Supporting Information:

- Supporting Information S1

Correspondence to:

S. J. Prichard,
sprich@uw.edu

Citation:

Prichard, S. J., Kennedy, M. C., Andreu, A. G., Eagle, P. C., French, N. H., & Billmire, M. (2019). Next-generation biomass mapping for regional emissions and carbon inventories: Incorporating uncertainty in wildland fuel characterization. *Journal of Geophysical Research: Biogeosciences*, 124, 3699–3716. <https://doi.org/10.1029/2019JG005083>

Received 8 FEB 2019

Accepted 16 OCT 2019

Accepted article online 4 DEC 2019

Published online 5 DEC 2019

Author Contributions:

Conceptualization: Susan J. Prichard, Maureen C. Kennedy, Paige C. Eagle, Nancy H. French, Michael Billmire

Formal analysis: Susan J. Prichard, Maureen C. Kennedy, Anne G. Andreu

Funding acquisition: Nancy H. French

Investigation: Susan J. Prichard, Maureen C. Kennedy

Methodology: Susan J. Prichard, Maureen C. Kennedy, Anne G. Andreu, Paige C. Eagle, Nancy H. French

Project administration: Nancy H. French

Resources: Nancy H. French

Software: Paige C. Eagle, Michael Billmire

(continued)

Next-Generation Biomass Mapping for Regional Emissions and Carbon Inventories: Incorporating Uncertainty in Wildland Fuel Characterization

Susan J. Prichard¹ , Maureen C. Kennedy² , Anne G. Andreu¹, Paige C. Eagle¹, Nancy H. French³ , and Michael Billmire³

¹School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA, ²Sciences and Mathematics, Division of the School of Interdisciplinary Arts and Sciences, University of Washington Tacoma, Tacoma, WA, USA,

³Michigan Tech Research Institute, Michigan Technological University, Ann Arbor, MI, USA

Abstract Biomass mapping is used in variety of applications including carbon assessments, emission inventories, and wildland fire and fuel planning. Single values are often applied to individual pixels to represent biomass of classified vegetation, but each biomass estimate has associated uncertainty that is generally not acknowledged nor quantified. In this study, we developed a geospatial database of wildland fuel biomass values to characterize the inherent variability within and across major vegetation types of the United States and Canada. For vegetation types that had sufficient quantification of biomass by fuel type (e.g., canopy, shrub, herbaceous, fine downed wood, coarse downed wood, and organic soil layers), we developed empirical distribution estimates. Based on available data, fitted distributions will be useful for informing the first-generation biomass mapping that incorporates variability in loading by vegetation and fuel type and to evaluate potential errors in point estimates given in current map products. Because combustible biomass is a common input in fire and smoke models, variability estimated for fitted distributions can be used to inform data input uncertainty in predictions of wildland fuel consumption and emissions and to provide stochastic inputs of biomass to ensemble simulation models.

Plain Language Summary The mass of live and dead vegetation, termed biomass, is used in variety of applications including carbon mapping, wildland smoke emission modeling, and fire management. In many mapping projects, single biomass values are often used to represent classified vegetation types. However, in reality, the biomass of grasslands, shrublands, and forests is extremely variable, and mapped values are associated with a high degree of uncertainty. In this study, we developed a database of wildland fuel biomass by major vegetation type in the United States and Canada. For well-studied vegetation types we developed mathematical models that statistically represent the distribution of biomass observations. Based on available data, fitted distributions will be useful for informing the first-generation biomass mapping that incorporates uncertainty estimates by major vegetation type to evaluate potential errors in current map products. Because biomass is a common input in fire and smoke models, fitted distributions can be used to evaluate the uncertainty in predicted wildfire emissions.

1. Introduction

Raster maps of vegetation and biomass are increasingly used in wildfire hazard assessments (Rollins, 2009; Scott et al., 2013), emission 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 (Peterson et al., 2018). 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 (e.g., Keane et al., 2013; McKenzie et al., 2007; Pierce et al., 2009; Riley et al., 2016; Rollins et al., 2004). Ideally, biomass estimates would be based on highly replicated and accurate field measurements to calibrate relationships with remotely sensed imagery and include estimated variability. In reality, biomass is composed of living vegetation (i.e., trees, shrubs, and herbs) and dead material (i.e., dead wood, litter, and organic soils) that is highly dynamic and variable across time and space (Keane et al., 2012). This inherent variability generally makes it impractical to collect enough measurements to represent the broad geographic and structural diversity of biomass in wildland environments.

Supervision: Susan J. Prichard, Maureen C. Kennedy, Nancy H. French
Validation: Susan J. Prichard, Maureen C. Kennedy, Anne G. Andreu, Paige C. Eagle
Visualization: Michael Billmire
Writing - original draft: Susan J. Prichard, Maureen C. Kennedy
Writing - review & editing: Anne G. Andreu, Nancy H. French

Biomass mapping generally relies on classifications based on mapped vegetation or interpretation of remotely sensed imagery rather than measured values (Keane et al., 2013). Many maps of combustible biomass, generally termed as wildland fuels for fire and smoke management, employ crosswalks from vegetation type (McKenzie et al., 2007; Raymond et al., 2006) to estimated biomass. LANDFIRE (<https://www.landfire.gov>) provides a central repository of mapped data for fuels and fire planning across the United States including surface and canopy fuel characteristics. 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 are derived from vegetation classifications from remote sensing imagery and other data layers (Peterson et al., 2015; Rollins, 2009). Wildland fuel maps generally include estimates of biomass per area by major fuel type (e.g., canopy, shrub, herbaceous, downed wood by size class, and organic soil layers including litter and duff). For emission inventories (e.g., McKenzie et al., 2012; Peterson et al., 2018; US EPA, 2017) biomass estimates are summarized at relatively coarse scales (1-km pixels) by aggregating finer-scale variability in wildland fuels (French et al., 2011).

Multispectral remote sensing imagery (e.g., Landsat) is useful for characterizing upper canopy fuels of dense forests and shrublands but is not valid for surface fuel characterization (e.g., understory vegetation, downed and dead wood, and organic soil layers). Advances are being made to map three-dimensional forest and shrubland canopy biomass through airborne and terrestrial light detection and ranging (lidar) data sets (Andersen et al., 2005; Skowronski et al., 2011), but these approaches are still under development and would offer only partial characterization of surface fuels. Because surface fuels are highly variable and are not easily predicted from canopy characteristics, there are generally high misclassification rates when canopy fuels are used to predict surface fuels (Jakubowski et al., 2013). In addition, the bulk of combustible biomass of surface fuels is within the first 10 cm of the fuelbed and often composed of intermixed live and dead vegetation that is difficult to quantify from remotely sensed imagery (Bright et al., 2016) and requires direct, ground-based measurement.

Uncertainty in biomass estimates underlies any wildland fuel mapping and is generally not acknowledged, much less quantified (Congalton et al., 2014; Urbanski, 2014). Regional assessments typically use point-based estimates as inputs to model applications and to inform management. However, whether the biomass is empirically estimated or modeled, estimates assigned to each mapped pixel have inherent sampling or prediction error (e.g., Krasnow et al., 2009; Thurner et al., 2013) due to the underlying variability in biomass. To avoid false precision in biomass estimation and resulting model predictions, it is necessary to quantify this variability in biomass for mapped values. When a mapped value is used as a data input into predictive models, this variability can then be used to inform plausible bounds on biomass input data. If the biomass value is not directly measured, then these plausible bounds represent a range of uncertainty for what biomass is actually on the ground—this allows the user to explore how this uncertainty is propagated through predictive models (see French et al., 2004).

Quantifying the variability in wildland fuels has important consequences for how biomass values are used in model applications. For models that require biomass as data input, it is important to understand how uncertainty in values propagates to uncertainty in model predictions (Hanna, 1988). There are multiple sources that contribute to the total uncertainty in model prediction, which can be classified broadly into four groups: model structure uncertainty, parameter estimation uncertainty, data input uncertainty, and natural variability over space and time (Beck, 1987; O'Neill & Gardner, 1979; Turley & Ford, 2009). Characterizing a pixel as a single point estimate masks the underlying variability in the data input, resulting in false precision in model predictions. A credible estimate of the variability in biomass associated with a given pixel or location requires knowledge of the likely statistical distribution of the input data. For example, with sufficient replication, field-based inventories of site estimates can be used to quantify an interval for total biomass. Absent direct measurement of biomass at a given time and space, such an interval can then be used to obtain a distribution of reasonable input values for a given model application. This would be an estimate of the uncertainty in the input value for the purposes of model prediction. A model prediction interval then can be obtained from the distribution of plausible input biomass values.

Evaluation of map accuracy in representing wildland fuels must also consider issues of scale and resolution. When we consider the quality of current continental-scale biomass, carbon, and wildland fuel maps, it is not appropriate to attempt to validate the estimate of an individual pixel against plot-level data. Plot-level data

Table 1
Data Sources Within the North American Wildland Fuel Database, With a Total of 26,620 Records

Name	Source	N
FIA loadings database	Forest Inventory and Analysis Program (2010), compiled and summarized by David Chojnacky, Virginia Tech University	13,964
Public LFRDB	The LANDFIRE Reference Data Base (LFRDB; LANDFIRE, 2012)	8,017
FLM database	Source data for the fuel loading model development (Lutes et al., 2009)	1,442
FOFEM fuels	Old database compiled to inform FOFEM fuel loading profiles (D. Lutes, personal communications)	917
Digital Photo Series database	Online database of natural fuels photo series sites (Wright & Eagle, 2007)	463
FFS database	Fire and fire surrogates (McIver et al., 2009)	339
Other	Published literature	1,478

are subject to sampling error and may not represent the pixel—such a validation will inevitably fail (Keane et al., 2013). Although point estimates represent our best estimate for an individual pixel, it does not match the on-the-ground value. Wildland fuels are variable at multiple spatial and temporal scales (Keane et al., 2012), and to improve the quality of our mapped products, we must harness this variability to understand the uncertainty in biomass estimation underlying each mapped point value. Note that for this study we focus on the component of data input uncertainty represented by the distribution of biomass values—we do not incorporate other sources of uncertainty such as the propagation of EVT classification error. This would likely add to total prediction uncertainty.

For wildland fuel applications, mapped biomass estimates have important consequences for smoke modeling and estimating emissions of air pollutants such as fine particulate matter ($PM < 2.5 \mu m$, $PM_{2.5}$). High concentrations of $PM_{2.5}$ have been linked to serious public health issues (Reisen et al., 2015); $PM_{2.5}$ emissions from wildfires are carefully monitored and are also commonly modeled for prescribed burn planning and to estimate wildfire smoke impacts (Liu et al., 2017). Errors in mapped biomass estimates that are either at the extreme low or high ranges of probable distributions can cascade into large errors in modeled smoke emissions (Larkin et al., 2012).

The primary goal of this study was to characterize the inherent variability in wildland fuels and facilitate uncertainty analysis in modeling applications. To do this, we developed the North American Wildland Fuel Database, a geospatial database of quantified biomass values over major vegetation types of the United States and Canada (<http://nawfd.mtri.org>). For vegetation types that had sufficient biomass records by fuel type we fit empirical distributions by fuel type. Published distributions will be useful for informing the first-generation biomass and wildland fuels mapping that incorporates uncertainty estimates by major category. Results of this study also will help inform future sampling needs to better represent the biomass of wildland fuels. Because biomass is a common input in fire and smoke models (Ottmar, 2014), data input intervals informed by our 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 emission inventories. As part of our analysis, we demonstrate sample distributions and how they can be used to evaluate mapped biomass estimates and modeled $PM_{2.5}$ emissions.

2. Approach and Methodology

The North American Wildland Fuel Database developed in this study stores existing dry-weight biomass observations by major fuel type across the United States. Our team began by compiling existing databases and importing wildland fuel biomass 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 (LFRDB; <https://www.landfire.gov/lfrdb.php>) were compilations of published literature and plot data (Table 1). We next conducted a literature review of biomass, fuel characterization and fuel consumption literature, and added observations from 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.

Table 2
Terms and Acronyms Use in This Paper

Term	Acronym	Definition
First Order Fire Effects Model	FOFEM	Software application used to predict fuel consumption, smoke emissions, soil heating, and tree mortality.
CONSUME	n/a	Software application used to predict fuel consumption and smoke emissions
Coarse woody debris	CWD	Coarse downed wood particles >7.6 cm in diameter.
Existing Vegetation Type group	EVT Group	Field within LANDFIRE (LINK) that classifies vegetation based on bioregional vegetation types at a broader grouping than Existing Vegetation Types.
Fuel Characteristic Classification System	FCCS	Software application used to summarize and classify wildland fuels, predict surface fire behavior, and calculate fuel loadings. FCCS fuelbeds are mapped within LANDFIRE.
Fuel Loading Model	FLM	Field within LANDFIRE that assigns a predicted surface fuel loading based on existing vegetation type and canopy layers.
Fuel type	n/a	Type of wildland fuel (canopy, shrub, herbaceous, fine downed wood, coarse downed wood, and organic soil layers) used to summarize biomass observations within the North American Wildland Fuel Database.

To group wildland fuel biomass observations, a standard mapping classification is needed. LANDFIRE is a widely used mapping source of geospatial fuels and vegetation containing over 640 existing vegetation types within 198 existing vegetation type groups (EVT Groups; Table 2). Given that the objective of the database was to quantify the distribution of fuel loads within vegetation types, we opted to use the more generic EVT group vegetation classification, provided within the LANDFIRE (2014) EVT layer, to ensure adequate sample sizes within each vegetation group. Using EVT group also reduced uncertainty in assigning vegetation type to each record compared to assignment by a more specific EVT. Most records within the database had 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 cross-walk tables to convert cover and forest types to EVT groups. For records that had only 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 were not georeferenced. 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 ($n = 2,470$) 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 EVT group. In many instances, simple summations were required to create summary inputs (e.g., herb load was calculated as the sum of forb and graminoid biomass and total coarse woody debris (CWD) is the sum of all sound and rotten coarse wood classes).

The database includes data from 271 sources from existing databases and scientific literature. Field definitions are provided in Table S1 in the supporting information. Entries from existing databases were presumed to be quality checked by the source agency and were not rechecked. As part of data entry and QA/QC, source references were carefully reviewed to ensure that they were not repeat values and not duplicative. We obtained the source reference and included a full citation for every record that had a published source reference. For quality assurance and quality control, we subsampled 30% of all source references and confirmed that entered data were accurate by cross checking entries with published values. Errors were uncommon; of records with data entry errors, most 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. We also flagged any extreme outliers in the database as observations in an EVT group (see section 2.5). These individual records were checked from the source data and any errors in recording were corrected. Otherwise the outliers were retained in the database.

As the database was compiled, supported database fields were expanded to accommodate varied studies and approaches. Many fuel 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.

2.1. Biomass Distributions

Database values were clustered by LANDFIRE EVT group for estimation of biomass distributions and included all observations across the entire geographic range of that EVT group. All analyses were conducted in the R statistical program (Version 3.4.1; R Core Team, 2017), and distributions were estimated using the R *fitdistrplus* package (Delignette-Muller & Dutang, 2015). To identify candidate distributions for individual fuel categories, we conducted an exploratory data analysis on select EVT groups with substantial representation (>1,000 entries). Histograms, box plots, and normal quantile-quantile 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 0, and the biomass distributions tended to be right-skewed rather than symmetric. Due to these features, we used hurdle estimation, described in the next section.

2.2. Hurdle Distribution Fitting

It is common in empirical studies of biomass (or, more commonly, abundance) for there to be high occurrence of zero values (Lecomte et al., 2013; Welsh et al., 1996). Often the nonzero part of the distribution is skewed to the right, implying that a distribution such as the lognormal or the gamma distribution is more appropriate than the normal distribution (Lecomte et al., 2013). One method to contend with high 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 nonzero values (Lachenbruch, 2002; Welsh et al., 1996). 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). By using this two-step procedure, we take advantage of robust maximum-likelihood estimation, which has to be approximated in a zero-inflated model. 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 nonzero value, and once that hurdle is crossed ($x > 0$) a continuous distribution is estimated for the data. The density function for the j^{th} fuel type in the k^{th} EVT group ($f_{kj}(x)$) can be written as (Lachenbruch, 2002)

$$f_{kj}(x, d) = \pi_{kj}^{1-d} \left((1 - \pi_{kj}) h_{kj}(x) \right)^d, \quad (1)$$

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

$$E(x) = (1 - \pi)E(h(x)) \quad (2)$$

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

$$h(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad x > 0 \quad (3)$$

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

$$E(x) = e^{\mu + \frac{\sigma^2}{2}} \quad (4)$$

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

$$h(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad x > 0 \quad (5)$$

with expected value

$$E(x) = \alpha\beta \quad (6)$$

Estimation of the hurdle distribution occurs in two steps. Let n_{kj} be the total number of entries in the database for a particular fuel type (j) in a particular EVT group (k), and x_{kji} be the i th value for fuel type j in EVT group k . Then

1. Estimate $\pi_{kj} = \frac{\sum I(x_{kji} = 0)}{n_{kj}}$ where I is an indicator function that takes a value of 1 if the entry has a value of 0, 0 otherwise and π_{kj} is the estimation probability of 0.
2. For the remaining nonzero 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 *nonzero 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 (cdf) at most 0.25 away from the true cumulative distribution (Massart, 1990). For example, if the theoretical cdf at some quantile X is 0.5, then we can estimate the empirical distribution within 0.25 of 0.5 at that value X .

2.3. 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.

2.3.1. 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 procedure to estimate the p value for the estimated distribution against the data (following Lilliefors, 1967). In general, a smaller p value (e.g., $p < \alpha$) gives evidence against the data following the estimated empirical distribution.

In the Monte Carlo procedure, we calculate the KS statistic for observed distribution relative to “theoretical” distribution at estimated parameter values. Then for 5,000 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 5,000 KS values when the null hypothesis is true; thus, a “null” distribution. The p value is then calculated as

$$1 - \frac{\sum_{i=1}^{n_{mc}} I(d_{obs} > d_i)}{n_{mc} + 1} \quad (7)$$

where n_{mc} is the number of MC replicates in the null distribution (5,000), d_i is the value of the KS statistic for Monte Carlo replicate i , and d_{obs} 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 that are smaller than the observed statistic. We divide by $n_{mc} + 1$ because we have $n_{mc} + 1$ total statistics (including d_{obs}). We can then evaluate, against some α level of significance ($\alpha = 0.05$), which distributions are “fail-to-reject.”

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 fail-to-reject 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 from 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.

2.3.2. Equivalence Tests

Robinson and Froese (2004) recommend an equivalence test to compare empirical data to model predictions using a two one-sided t test. 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; ϵ), then the null hypothesis is rejected and the observed data are judged to be “equivalent” to the theoretical distribution (within the error tolerance). Therefore, equivalence is in the alternative rather than the null hypothesis. Here we use t tests to assess adequate matching between our observed empirical cumulative distribution of fuel type and the theoretical cumulative distribution function associated with each candidate distribution. Let $x_{(i)}$ be the i th quantile of the empirical data distribution, and $x_{(i)}$ be the i th quantile of the theoretical distribution. Then the difference between the observed and theoretical cumulative distributions (x_{di}) is

$$x_{di} = x_{(i)} - x_{(i)} \quad (8)$$

We then calculate x_{di} as the mean distance between observed and theoretical cumulative distributions and use t tests 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) used 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 are considered fuel types with a KS p value $>\alpha$, and a small ϵ value for the equivalence procedure outlined above. We assign broad goodness of fit classifications based on these two goodness of fit metrics (Table S2) in combination with the sample size. A fit is considered excellent if it is based on ≥ 100 entries, associated with a nonsignificant KS p value, and an ϵ value ≤ 0.05 . A fit is considered good if it is based on ≥ 30 entries, has a nonsignificant KS p value, and $0.05 < \epsilon \leq 0.15$; alternatively, a fit is considered good if it has >30 entries, and a significant KS p value associated with an $\epsilon \leq 0.05$. A fit is considered poor if it has ≥ 30 entries, associated with a significant KS p value and a large (>0.15) ϵ value. The distribution is not estimated for any fuel type X EVT group combination with <30 entries, and assigned an NA here.

2.4. 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 5,000 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.

2.5. Outliers

As part of our quality assurance effort, we identified extreme outliers in the database as any value $>Q_3 + 4 * \text{IQR}$, where Q_3 is the third quartile for the empirical distribution and IQR is the interquartile range ($Q_3 - Q_1$). 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 extreme outliers. All distribution estimates, both with and without outliers, are presented in the supporting information.

Table 3

Percentage of Total U.S. 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

2.6. Sample Distributions

For the purposes of demonstrating comparisons of distributions among fuel types and EVT groups we present contrasting examples of EVT groups: 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). These were chosen because they represent major EVT groups across the continental United States and because they also had sufficient representation of the chosen fuel types to estimate distributions. For these EVT groups we present biomass distributions for total aboveground biomass of trees, coarse woody debris (downed wood

>7.6-cm diameter), duff, and litter.

To place the biomass distributions in the context of mapped biomass values for smoke management and emission estimation, we made use of existing mapped surface fuel type values (CWD, litter, and duff) within LANDFIRE (LANDFIRE, 2008) and compared them to fitted distributions for three common existing vegetation type groups in the eastern United States and three in the western United States and Alaska. These EVT groups were selected based on having good to excellent distribution fits for CWD, litter, and duff. From biomass estimates, we used Consume version 5.0 algorithms (Prichard et al., in prep) to calculate fuel consumption and emissions of fine particulate matter less than 2.5 μm ($\text{PM}_{2.5}$). Mapped values were obtained for sample study areas for each EVT group (Table S3) by selecting an area that was dominated by a single 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 Fuel Characteristic Classification System (FCCS) fuelbed and fuel loading model (FLM) that was crosswalked to that EVT group. Emissions were modeled in Consume using a dry fuel moisture to represent high availability of fuels for burning (15% for CWD, 11% for litter, and 30% for duff). We then compared the mapped FCCS and FLM estimates and associated emission values to a 75% prediction interval using the distribution estimate from the database.

3. Results

Of the 198 LANDFIRE (2014) EVT groups, the database contains records for 134 EVT groups. A total of 68 EVT groups had sufficient entries to estimate at least one fitted distribution of a fuel type (Table S3). Based on broad physiognomic or land use category, the largest percentage of land area in the United States is classified as forest and woodland (32%) followed by shrublands (19%), agriculture (17%), and nonvegetated pixels (16%). However, the percentage of EVT groups with sufficient record counts for distribution fitting is overrepresented by forest and woodland EVT groups (70%) with only 22% and 13% of shrubland and grassland EVT groups represented by at least one fitted distribution (Table 3). Of the forest and woodland EVT groups there was higher representation of coniferous forests (78%) than broadleaf forests (68%) and mixed forests (63%) for the fitted distributions. Mapped locations of records within the database reveal much higher record availability in forested regions of the United States and fewer records for nonforested regions, including much of the Central United States (Figure 1).

3.1. Empirical Distributions

To evaluate the ranges of observed biomass in different ecosystems, we present six representative EVT groups (three for eastern hardwood and three for mixed conifer) and fuel types (Figure 2). The example fits presented are all classified as “good” or “excellent.” A common characteristic of many distributions is the high proportion of zero values and high coefficients of variation (CV). The results of all estimated distributions are given in Table S3 in the supporting information.

The empirical distributions of biomass compare well to estimated density functions (Figures 3 and 4). Note that in Figures 3 and 4 the histograms give the empirical distribution, with the associated estimated density function overlaid. The density axis is scaled to the histogram bin width so that the total area across each histogram is 1. In Figures 3 and 4 the location of the distribution mode as well as the density in the tail is well-represented by the density function.

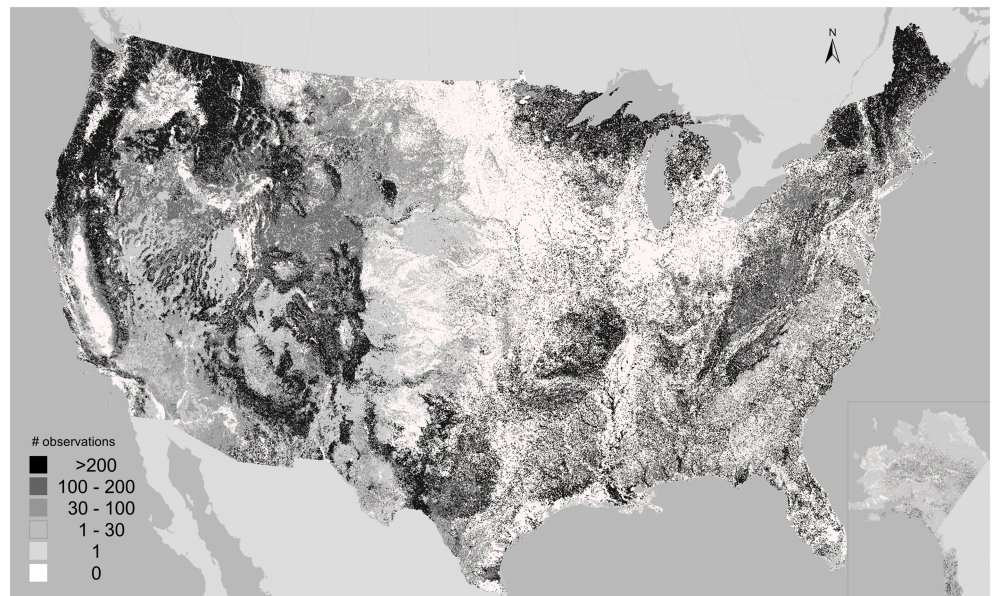


Figure 1. Mapped locations of fuel loading records grouped by number of observations in the conterminous United States and Alaska. An online, dynamic version of this map is at <http://nawfd.mtri.org>.

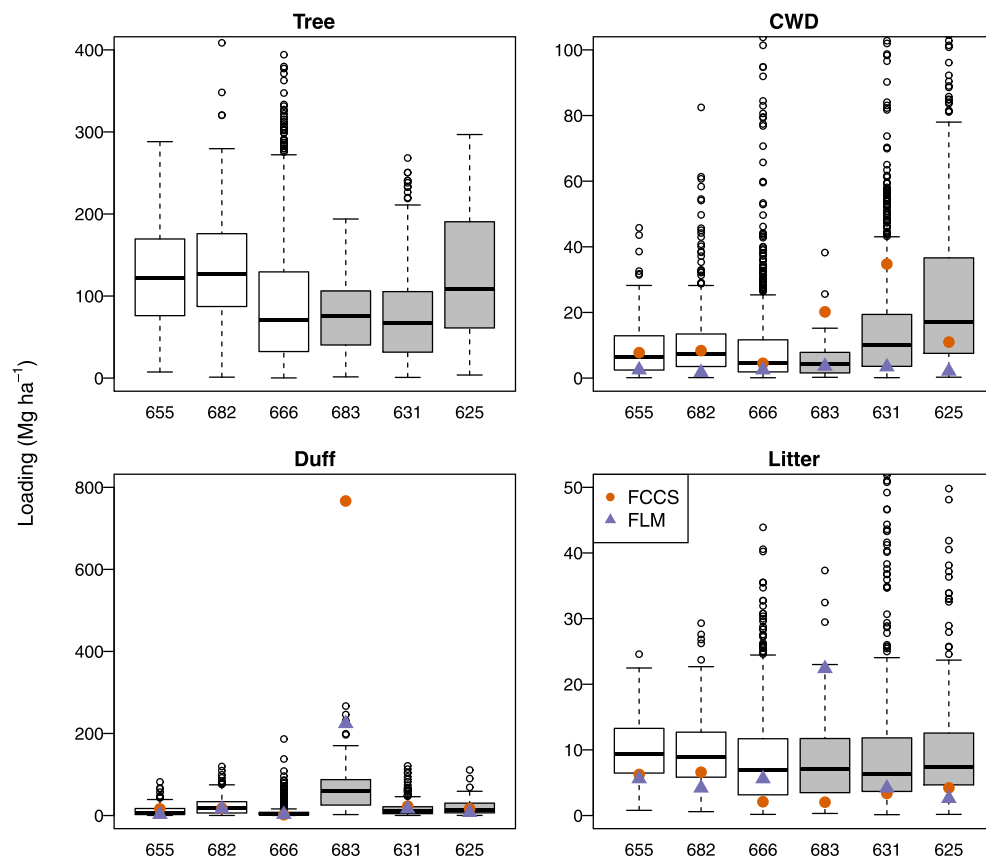


Figure 2. Box plot comparison fuel loading ranges by Existing Vegetation Type (EVT) group. White boxes are hardwood forests, grey boxes conifer forests. For some fuel categories, distributions are very similar (e.g., trees and litter), and for others there are some differences in distributions (CWD, duff). There is high variability regardless of EVT group. EVT groups include 655 = beech-maple-basswood forest; 682 = yellow birch-sugar maple forest; 666 = eastern floodplain forest; 683 = peatland forest; 631 = ponderosa pine forest and woodland; 625 = Douglas-fir, ponderosa pine, and lodgepole pine forest. Points give mapped values from sample areas within the LANDFIRE, 2008 Fuel Characteristic Classification System (FCCS) and Fuel Loading Model (FLM) layers.

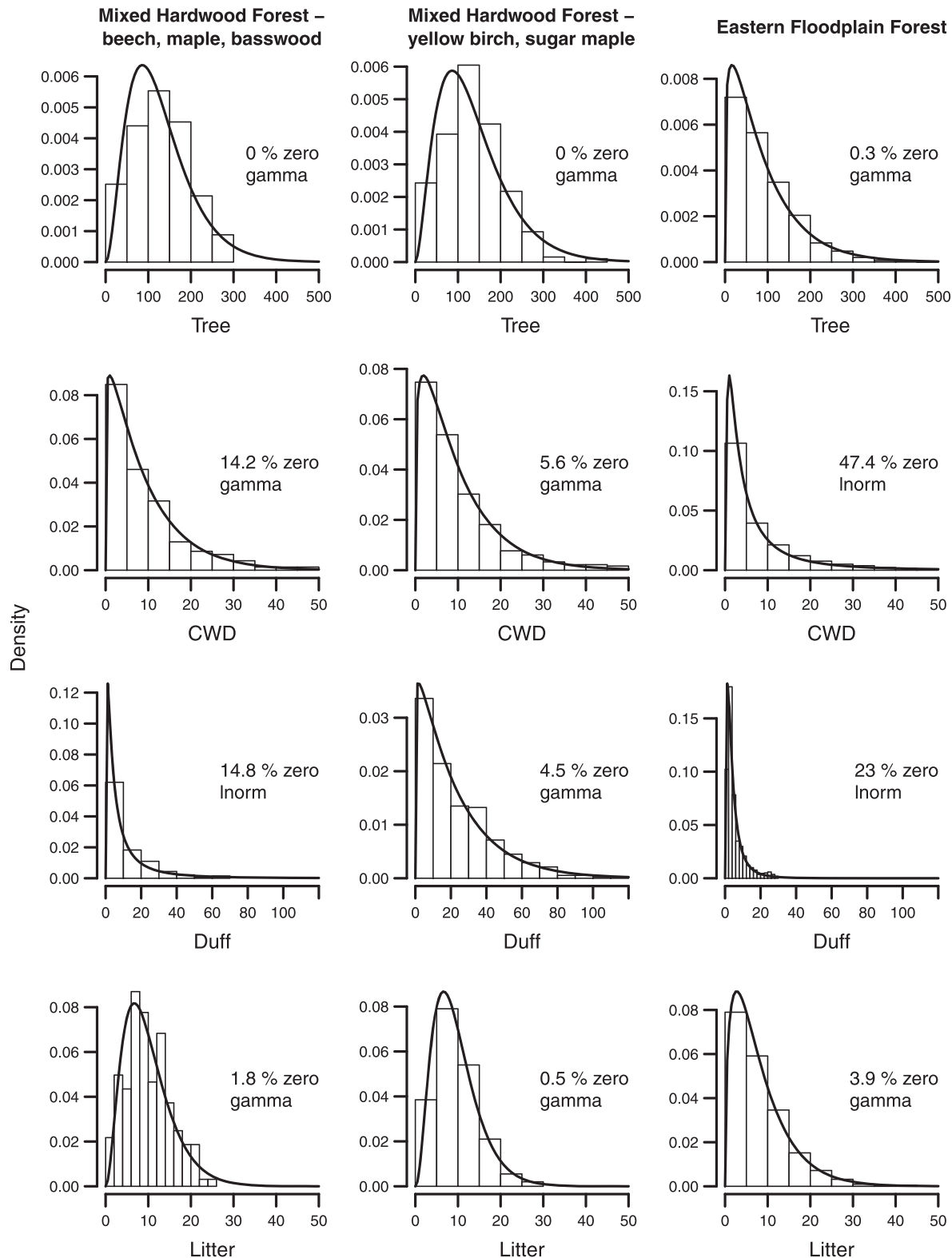


Figure 3. Empirical (histogram) and estimated density function (solid line) for biomass distributions (Mg/ha) for aboveground trees (Tree), coarse woody debris (CWD), and organic soil layers (Duff and Litter) of three sample eastern mixed hardwood Existing Vegetation Type groups. Note that histogram bin widths are determined by the default R setting—for heavily skewed data with large outliers this results in fewer bins of larger width. The density y axis is scaled so that the total area in the histogram is one, and is not comparable between graphs. The estimated density function is scaled so it matches the histogram scaling. Graphs are adapted from the denscomp function in the fitdistr R package.

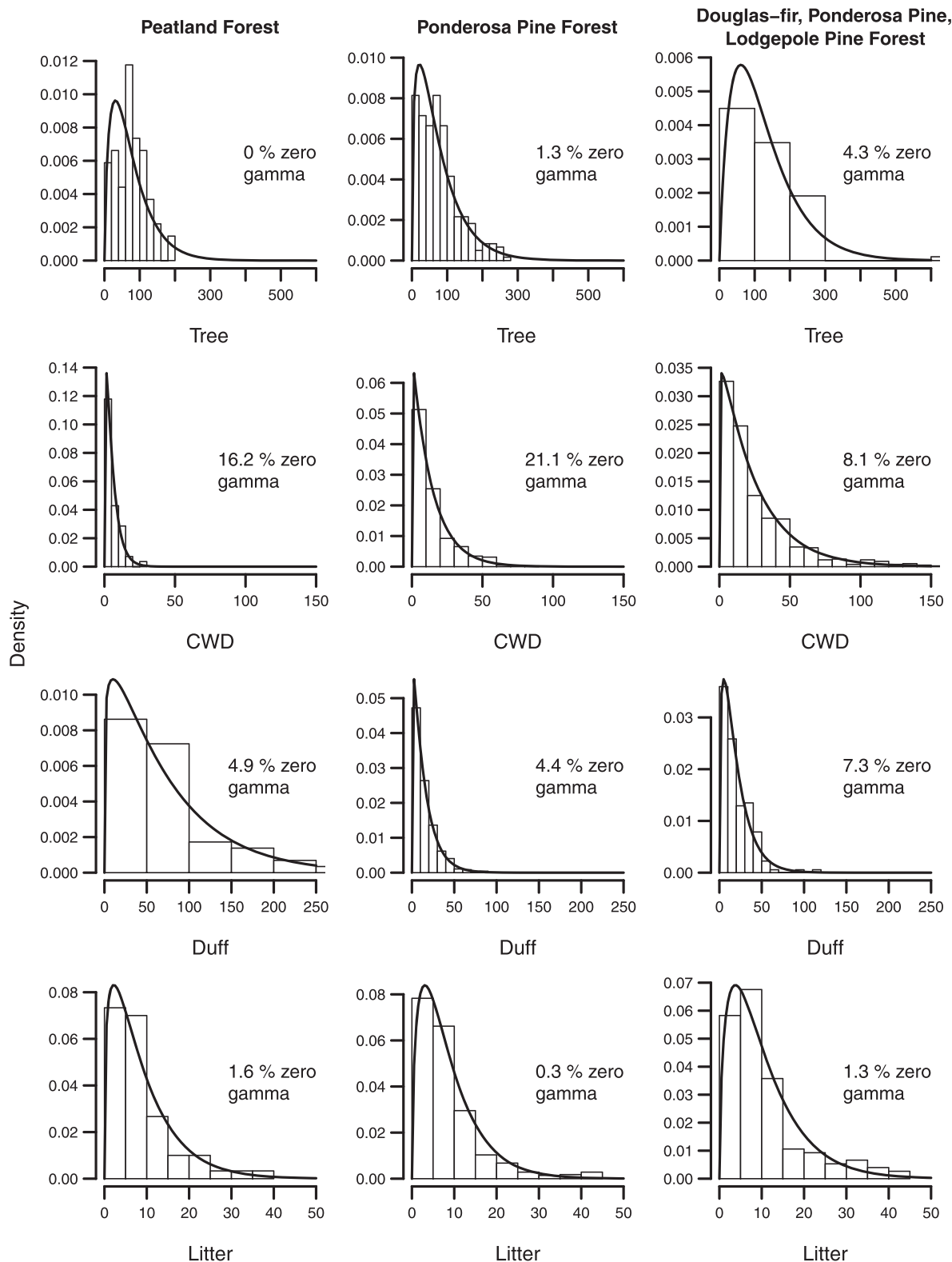


Figure 4. Empirical (histogram) and estimated density function (solid line) for biomass distributions (Mg/ha) for aboveground trees (Tree), coarse woody debris (CWD), and organic soil layers (Duff and Litter) of three sample western conifer Existing Vegetation Type groups. Extreme outliers are excluded from the database distributions. Note that histogram bin widths are determined by the default R setting—for heavily skewed data with large outliers this results in fewer bins of larger width. The density y axis is scaled so that the total area in the histogram is one, and is not comparable between graphs. The estimated density function is scaled so it matches the histogram scaling. Graphs are adapted from the denscomp function in the fitdistr R package.

Table 4
Empirical Distribution Summaries for Example Fuel Types and Existing Vegetation Type (EVT) Groups

ID	Group name	Fuel	Prop 0	n	Q ₁	Q ₂	Q ₃	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	1,513	32.34	71.3	129.44	0.86
		CWD	0.474	813	1.89	4.69	11.66	1.63
		Duff	0.23	1,164	2.3	3.45	7.75	1.7
		Litter	0.039	1,453	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

Note. 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 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.

For common forest types in the continental United States, total aboveground biomass distributions are best represented by a gamma distribution (with nearly no zero values). Observed values range from near 0 to near 500 Mg/ha, with variability depending on EVT group (Table S4). Median tree total aboveground biomass is comparable for mixed hardwood sites, ranging from 71 Mg/ha in eastern floodplain forests to just above 120 Mg/ha in yellow birch and sugar maple (YB-SM) forests and high standard deviations and coefficients of variation. Tree biomass ranges from 0 to 500 Mg/ha for conifer forests with highest median (Q₂) biomass value in Douglas-fir/ponderosa pine/lodgepole pine (DF-PP-LP) forests (108 Mg/ha) compared to 75 Mg/ha for the other forest types.

Biomass of CWD is particularly variable across records and contributes to high coefficients of variation in distribution estimates. For example, in the three sample mixed hardwood forest distributions (Figure 3), CWD ranges from 0 to 50 Mg/ha (Table 4), and the proportion of zero values is 5.6% in YB-SM distributions, 14% in beech-maple-basswood forests, and nearly half of all records for eastern floodplain forests (47%). Median CWD biomass values are quite similar across all mixed hardwood sites as are the shapes of the distributions. In the three conifer forest distributions highlighted here (Figure 4), CWD ranges from 0 to nearly 150 Mg/ha, and the proportion of records with a zero value ranges from 8% to 21%. Median CWD is greatest in DF-PP-LP forests and lowest in peat forests. In each case, standard deviations meet or exceed estimated median values.

Duff biomass also has relatively high variability, and records have a relatively low proportion of zero values (4.4% to 7.3%) compared to CWD. Presence of duff is more variable in the three sample mixed hardwood

forests than conifer forests with the proportion of records without duff ranging from 4.5% to 23%. Median duff biomass ranges from 3 to 17 Mg/ha in mixed hardwood EVT groups, and standard deviations either exceed or are close to the estimated median value. Duff biomass ranges widely from 0 to 250 Mg/ha in peatland forests and 0 to >100 Mg/ha in the other sample conifer forests. Median duff biomass is markedly higher in peatland forests (59 Mg/ha) compared to the other conifer forests (11–15 Mg/ha), and as with CWD, standard deviations are high.

Litter is consistently present across the six forest types with less than 5% of records with zero values in the database records. Median values are markedly similar in eastern hardwood forests (7–9 Mg/ha), with moderate to high coefficients of variation. In mixed conifer forests, median values are comparable (6–7 Mg/ha) and have slightly higher CV than eastern hardwood forests.

3.2. Comparison With Mapped Estimates

Mapped biomass values from FCCS and FLMs across the six representative forest types often fall within fitted distributions but with notable exceptions. For CWD, FCCS estimates match median distribution fits for eastern hardwoods but are high for peatland forests, exceeding the Q3 value, and are also high for ponderosa pine forests (Figure 2). In contrast, the FCCS CWD estimate for Douglas-fir-ponderosa pine-lodgepole pine 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 PM_{2.5} emissions under a dry fuel moisture scenario, mapped FCCS values produce similar emission estimates to the median (Q2) distribution fit for the eastern hardwoods but are high for peatland and ponderosa pine forests and low for Douglas-fir-ponderosa pine-lodgepole pine forests (Figure 5). The FLM CWD estimate generally underpredicts emissions relative to fitted distributions. For example, for the Douglas-fir-ponderosa pine-lodgepole pine EVT group, using the mapped FCCS value to estimate emissions would predict 292 versus 78 kg/ha from the FLM mapped value and 696 kg/ha for the median fitted distribution (Q2) value. Mapped FCCS and FLM estimates of litter generally are much lower than Q1 values for the six representative forest types. However, the FLM estimate for peatland litter 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 PM_{2.5} emissions for these forest types. Eastern hardwood forests have little duff, reflected in fitted distribution values, and mapped FCCS and FLM values. Mapped FCCS and FLM biomass values for ponderosa pine and Douglas-fir-ponderosa pine-lodgepole pine forests are both well within the range of distributions. However, the FCCS value for peatland forests is nearly 4 times that of the Q3 loading value. The FLM value also exceeds the Q3 value but is still within the fitted distribution. Because duff emission factors are much higher than for CWD and litter, the consequences of errors in biomass estimates are also greater. This is revealed for the peatland forests with even greater disparities in PM_{2.5} emissions than in the actual biomass values.

4. Discussion

Wildland fuel information and maps are used as inputs to numerous models, including consumption and emission models such as the First Order Fire Effects Model (Reinhardt et al., 1997) and Consume (Ottmar, 2014; Prichard et al., 2014), wildland fire behavior prediction tools such as FLAMMAP (Finney, 2006), and smoke-dispersion models and frameworks such as the Wildland Fire Emissions Information System (French et al., 2014) and BlueSky (Larkin et al., 2014). For many modeling studies of biomass, carbon fluxes, 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. However, to date, distributions of mapped fuels have not been available for simulating fire behavior, fire 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 variability in estimating the biomass of wildland fuels in North America.

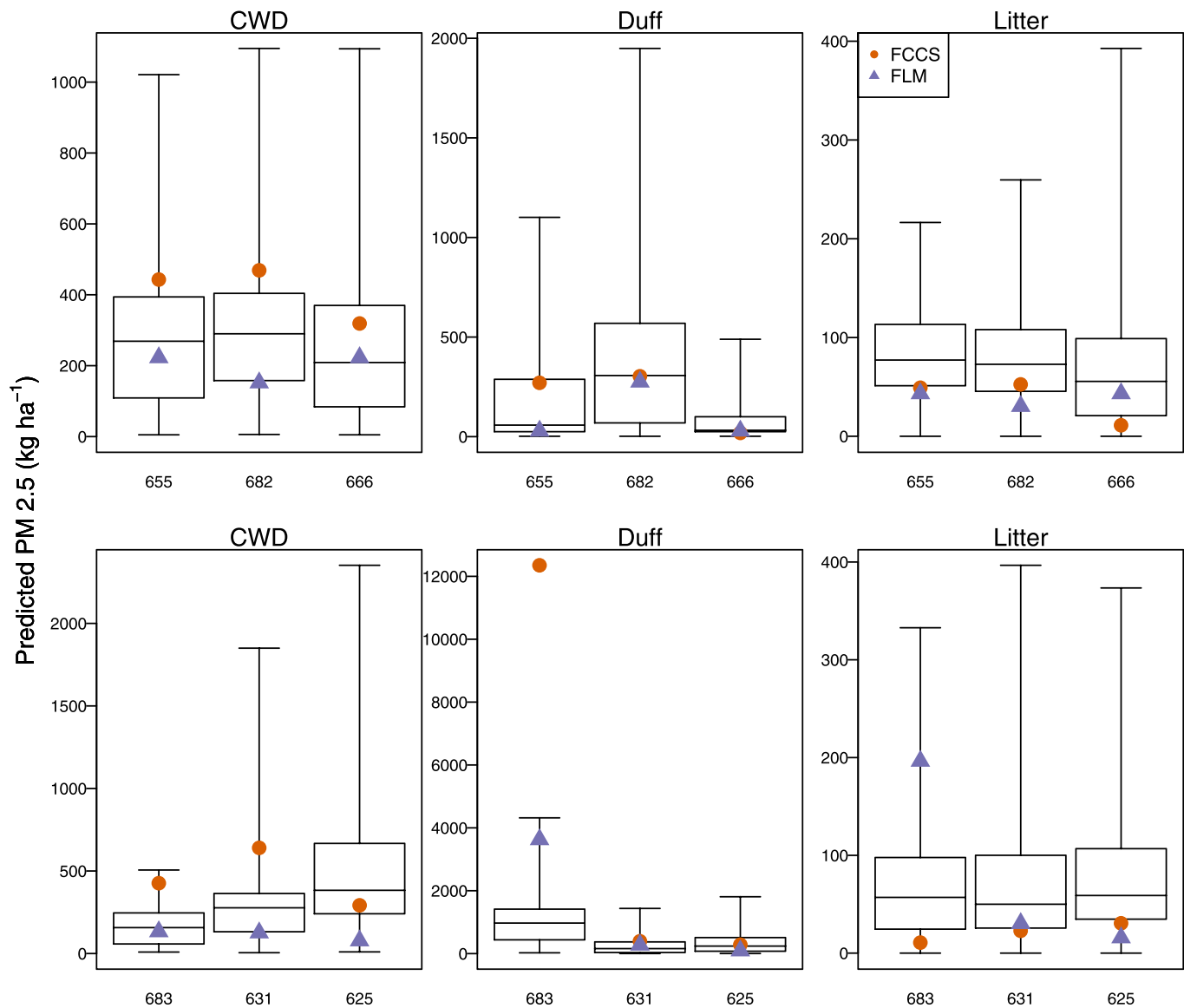


Figure 5. Boxplot of calculated $PM_{2.5}$ emissions compared between database distributions and mapped values for that Existing Vegetation Type (EVT) group, obtained from sample areas within the LANDFIRE, 2008 Fuel Characteristic Classification System (FCCS) and Fuel Loading Model (FLM) layers. Extreme outliers are excluded from the database distributions. EVT groups include 655 = beech-maple-basswood forest; 682 = yellow birch-sugar maple forest; 666 = eastern floodplain forest; 683 = peatland forest; 631 = ponderosa pine forest and woodland; 625 = Douglas-fir, ponderosa pine, and lodgepole pine forest.

The database offers a compilation of existing data sources on surface and canopy fuels for the United States and Canada and is designed to be expandable to include values from Mexico and new studies. As expected, we observed a wide range in biomass by major fuel type (aboveground tree biomass, shrubs, herbs, CWD, litter, and duff). Distributions were generally best fitted with either a gamma or lognormal distribution. Of the 134 EVT groups supported in the database, we are able to estimate at least one fitted distribution for 64 EVT groups (48%). Fitted distributions can be used to provide stochastic draws from the distributions or stratified samples across percentiles, including the 25th, 50th, and 75th percentile values presented in this paper.

Because biomass and carbon are most commonly mapped using crosswalks between a vegetation classification and assigned values, we chose to classify biomass observations by vegetation type, using the LANDFIRE EVT Group layer. However, the observed variability is so high within EVT groups that it suggests that an even broader grouping variable by biophysical setting and physiognomic type might be warranted. Based

on this finding, we are planning to add a broader classification to the database that will contribute to coarser scale biomass mapping. Database development also revealed substantial bias in data collection toward forest types versus nonforest types (Table 3). 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 data sets for informing uncertainty in carbon mapping and emission inventories, the data gaps revealed in this study justify field-based fuel characterization in nonforest vegetation types and inform future sampling. 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 five most common shrubland types, we had sufficient records 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 database. Agricultural EVT groups constitute a large percentage of U.S. land area and are not currently represented in the database. In particular, for biomass and carbon mapping applications, records are needed to represent pasturelands, row crops, and wheat fields. Of the major forest types, plantations and white spruce forests and woodlands lack sufficient records to support fitted distributions. Another issue with low representation of many EVT groups is the lack of specification in vegetation types and associated descriptive data. Of the six 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 refine 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 most likely vegetation.

It is also possible that there are smaller-scale sources of variability in measured biomass such as local abiotic conditions. Evaluating these sources of variability is beyond the scope of this current study. Future sampling efforts should also identify representative local abiotic conditions for a given grouping so any potential bias in those conditions can be identified.

Even with substantial data gaps identified in this study, the fuel loading data set should be immediately useful for applications including carbon accounting, fire hazard assessments, and emission inventories. The database has been constructed to facilitate data entry as additional data sets become available, making this a “living” database. As such, it offers guidance on where additional field inventories and fuel characterization would be particularly beneficial. Fitted biomass distributions and the underlying database currently allow for three major advances in fuel characterization:

- By using distributions of fuel loading for a vegetation type rather than a point-based map estimate, a credible interval of emission estimates for carbon accounting can be generated. The ability to calculate uncertainty bounds on model predictions provides users with a plausible range of model outputs rather than a single point estimate.
- 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 to determine if the mapped values are representative of known EVT distributions. For example, mapped values based on FCCS and FLM layers within LANDFIRE generally fall within the observed distribution of the total aboveground biomass of trees, CWD, litter, and duff. This finding confirms that the pixel-based estimates of biomass are within the observed range and would provide a reasonable point estimate of fuel consumption and $PM_{2.5}$ emissions. However, the FCCS mapped duff value for peatland forests is clearly not in the center of the distribution estimated here and has large consequences for predicted $PM_{2.5}$ emissions (Figure 4). 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 inform other improvements in mapped products.
- Providing distribution fits by major fuel type 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 propagates to uncertainty in wildfire emissions. For a given EVT group, a sensitivity analysis of fuel-loading inputs to emission 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.

5. Conclusions

This study presents the first uncertainty estimates in biomass inputs commonly used in carbon inventories, fire hazard analysis, and emission estimation. The use of point estimates of wildland fuel biomass for model applications is problematic. Individual point estimates are unlikely to be absolutely correct, but they represent the best value for a given pixel at a given time. 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 data sets there is tremendous variability in biomass values, which propagates to uncertainty in associated model applications. Model predictions that use point estimates of wildland fuel biomass are therefore vastly underrepresenting the uncertainty associated with those model predictions.

Even with the large number of records in the North American Wildland Fuel Database ($n = 26,620$), there are 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 EVT but quickly realized there was insufficient coverage of observations to estimate distributions by vegetation type. This led us to aggregate by EVT group, but we find 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 prediction uncertainty.

The database is a work in progress and will benefit from additional records, particularly for underrepresented fuel types. Fitted distributions for 65 EVT Groups will be useful to assessing how representative existing mapped estimates of fuel loadings are, calculating variability in fuel-loading estimates, and providing estimates of potential wildland fire emissions and carbon stores and associated uncertainty.

Acknowledgments

We acknowledge funding from the Joint Fire Science Program under Project JFSP 15-1-01-1 and the U.S. Forest Service Pacific Northwest Research Station. We are grateful to Duncan Lutes and David Chojnacky for assistance in data compilation and Danielle Tanzer for the assistance with data entry and quality assurance. Don McKenzie contributed to project design and offered constructive comments on an earlier draft of this paper. Marion LaRocque assisted in developing R scripts for distribution fitting. The North American Wildland Fuel Database is archived with the U.S. Forest Service Research Data Archive at Prichard, J., Kennedy, M. C., Andreu, A. G., Eagle, P. C., and French, N. H. 2019. North American Wildland Fuel Database. Fort Collins, CO: Forest Service Research Data Archive (<https://doi.org/10.2737/RDS-2019-0025>).

References

- Andersen, H. E., McGaughey, R. J., & Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment*, 94, 4414–4449. <https://doi.org/10.1016/j.rse.2004.10.013>
- Anderson, H. E. (1982). *Aids to determining fuel models for estimating fire behavior*. INT-GTR-122. Ogden, UT: US Forest Service Intermountain Forest and Range Experiment Station.
- Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research*, 23, 1393–1442. <https://doi.org/10.1029/WR023i008p01393>
- Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., et al. (2007). Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, 112(4), 1658–1677. <https://doi.org/10.1016/j.rse.2007.08.021>
- Bright, B. C., Loudermilk, E. L., Pokswinski, S. M., Hudak, A. T., & O'Brien, J. J. (2016). Introducing close-range photogrammetry for characterizing forest understory plant diversity and surface fuel structure at fine scales. *Canadian Journal of Remote Sensing*, 42, 460–472. <https://doi.org/10.1080/07038992.2016.1229598>
- Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., & Ozdogan, M. (2014). Global land cover mapping: A review and uncertainty analysis. *Remote Sensing*, 6, 12,070–12,093. <https://doi.org/10.3390/rs61212070>
- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64, 1–34. <http://www.jstatsoft.org/v64/i04>
- Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A Comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, 1, 115–126. <https://doi.org/10.2307/1391852>
- 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, pp. 213–220)*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Forest Inventory and Analysis Program (2010) <http://web.gis.vt.edu/forestry/dwm> [Accessed 6 February 2019]
- French, N. H. F., de Groot, W. J., Jenkins, L. K., Rogers, B. M., Alvarado, E., Amiro, B., et al. (2011). Model comparisons for estimating carbon emissions from North American wildland fire. *Journal of Geophysical Research*, 116, G00K05. <https://doi.org/10.1029/2010JG001469>
- French, N. H. F., Goovaerts, P., & Kasischke, E. S. (2004). Uncertainty in estimating carbon emissions from boreal forest fires. *Journal of Geophysical Research*, 109, D14S08. <https://doi.org/10.1029/2003JD003635>
- French, N. H. F., McKenzie, D., Erickson, T., Koziol, B., Billmire, M., Endsley, K. A., et al. (2014). Modeling regional-scale fire emissions with the Wildland Fire Emissions Information System. *Earth Interactions*, 18, 1–26. <https://doi.org/10.1175/EI-D-14-0002.1>
- Gregg, T. F., & Hummel, S. (2002). Assessing sampling uncertainty in FVS projections using a bootstrap resampling method. *USDA Forest Service Proceedings 2002 In Proceedings of the Second Forest Vegetation Simulator Conference, 2002 February 12–14 (RMRS-P-25)*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Hanna, S. R. (1988). Air quality model evaluation and uncertainty. *Journal of the Air Pollution Control Association*, 38, 406–412. <https://doi.org/10.1080/08940630.1988.10466390>
- Jakubowski, M. K., Li, W., Guo, Q., & Kelly, M. (2013). Delineating individual trees lidar data: a comparison of vector- and raster-based segmentation approaches. *Remote Sensing*, 5, 4163–4186.

- 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. <https://doi.org/10.1071/WF11139>
- Keane, R. E., Gray, K., Baccio, V., & Leirfallom, S. (2012). Spatial scaling of wildland fuels for six forest and rangeland ecosystems of the northern Rocky Mountains, USA. *Landscape Ecology*, 27, 1213–1234. <https://doi.org/10.1007/s10980-012-9773-9>
- Keane, R. E., Herynk, J. M., Toney, C., Urbanski, S. P., Lutes, D. C., & 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. <https://doi.org/10.1016/j.foreco.2013.06.001>
- Krasnow, K., Schoennagel, T., & 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. <https://doi.org/10.1016/j.foreco.2009.01.020>
- Lachenbruch, P. A. (2002). Analysis of data with excess zeros. *Statistical Methods in Medical Research*, 11, 297–302. <https://doi.org/10.1191/0962280202sm289ra>
- LANDFIRE (2008). Fuel loading models and Fuel Characteristics Classification System layers, LANDFIRE 1.1.0, U.S. Department of the Interior, Geological Survey. [Accessed 7 February 2019 at <http://landfire.cr.usgs.gov/viewer>]
- LANDFIRE (2012). LANDFIRE research database. <https://www.landfire.gov/DataDictionary/LANDFIRELFRDBDataDictionary.pdf> [Accessed 7 February 2019 at <http://landfire.cr.usgs.gov/viewer>]
- LANDFIRE (2014). Existing vegetation type layer, U.S. Department of the Interior, Geological Survey. [Accessed 7 February 2019 at <https://www.landfire.gov/evt.php>]
- Larkin, N. K., Raffuse, S. M., & Strand, T. M. (2014). Wildland fire emissions, carbon, and climate: US emissions inventories. *Forest Ecology and Management*, 317, 61–69. <https://doi.org/10.1016/j.foreco.2013.09.012>
- Larkin, N. K., Stand, T. T., Drury, S. A., Raffuse, S. M., Solomon, R. C., O'Neill, S. M., et al. (2012). Phase 1 of the Smoke and Emissions Model Intercomparison Project (SEMI-P): Test cases, methods, and analysis results; Joint Fire Science Program: Boise, ID, USA. Vol. 42.
- Lecomte, J. B., Benoît, H. P., Ancelet, S., Etienne, M. P., Bel, L., & Parent, E. (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. <https://doi.org/10.1111/2041-210X.12122>
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399–402. <https://doi.org/10.2307/2283970>
- Liu, X., Huey, L. G., Yokelson, R. J., Selimovic, V., Simpson, I. J., Müller, M., et al. (2017). Airborne measurements of western U.S. wildfire emissions: Comparison with prescribed burning and air quality implications. *Journal of Geophysical Research: Atmospheres*, 122, 6108–6129. <https://doi.org/10.1002/2016JD026315>
- Lutes, D. C., Keane, R. E., & Caratti, J. (2009). A surface fuel classification for estimating fire effects. *International Journal of Wildland Fire*, 18, 802–814. <https://doi.org/10.1071/WF08062>
- Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18, 1269–1283.
- McIver, J. D., Youngblood, A. P., & Stephens, S. L. (2009). The national Fire and Fire Surrogate study: ecological consequences of fuel reduction methods in seasonally dry forests. *Ecological Applications*, 19(2), 283–284. [Accessed 7 February 2019 at <http://www.fs.usda.gov/rds/archive/Product/RDS-2016-0009>, <https://doi.org/10.1890/07-1785.1>]
- McKenzie, D., French, N. H., & Ottmar, R. D. (2012). National database for calculating fuel available to wildfires. *Eos, Transactions American Geophysical Union*, 93(6), 57–58. <https://doi.org/10.1029/2012EO060002>
- McKenzie, D., Raymond, C. L., Kellogg, L.-K. B., Norheim, R. A., Andreu, A. G., Bayard, A. C., et al. (2007). Mapping fuels at multiple scales: landscape application of the Fuel Characteristic Classification System. *Canadian Journal of Forest Research*, 37, 2421–2437. <https://doi.org/10.1139/X07-056>
- O'Neill, R. V., & Gardner, R. H. (1979). Sources of uncertainty in ecological models. In B. P. Zeigler, M. S. Elzas, G. J. Klir, & T. I. Oren (Eds.), *Methodology in Systems Modelling and Simulation* (pp. 447–463). Amsterdam, New York, Oxford, New York, US: North-Holland Publishing Company.
- Ottmar, R. D. (2014). Wildland fire emissions, carbon, and climate: Modeling fuel consumption. *Forest Ecology and Management*, 317, 41–50. <https://doi.org/10.1016/j.foreco.2013.06.010>
- Ottmar, R. D., Sandberg, D. V., Riccardi, C. L., & 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. <https://doi.org/10.1139/X07-077>
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A large and persistent carbon sink in the world's forests. *Science*, 333, 988–993.
- Peterson, B., Nelson, K. J., Seielstad, C., Stoker, J., Jolly, W. M., & Parsons, R. (2015). Automated integration of lidar into the LANDFIRE product suite. *Remote Sensing Letters*, 6, 247–256. <https://doi.org/10.1080/2150704X.2015.1029086>
- Peterson, J., Lahm, P., Fitch, M., George, M., Haddow, D., Melvin, M., et al. (2018). *NWCG smoke management guide for prescribed fire (PMS 420-2, NFES 001279)*. Boise, Idaho: National Wildfire Coordinating Group.
- Pierce, K. B. Jr., Ohmann, J. L., Wimberly, M. C., Gregory, M. J., & 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. <https://doi.org/10.1139/X09-102>
- Prichard, S. J., Karau, E. C., Ottmar, R. D., Kennedy, M. C., Cronan, J. B., Wright, C. S., & Keane, R. E. (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*, 44, 784–795. <https://doi.org/10.1139/cjfr-2013-0499>
- Quillet, A., Peng, C., & Garneau, A. (2010). Toward dynamic global vegetation models for simulating vegetation-climate interactions and feedbacks: Recent developments, limitations, and future challenges. *Environmental Reviews*, 18, 333–353.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raymond, C. L., Kellogg, L. K. B., & McKenzie, D. (2006). *Mapping Fuels on the Okanogan and Wenatchee National Forests. In: Fuels management—How to measure success: Conference Proceedings. 28-30 March 2006 (RMRS-P-41)*. Ogden, Utah: U.S. Forest Service Rocky Mountain Research Station.
- Reinhardt, E. D., Keane, R. E., & Brown, J. K. (1997). *First order fire effects model: FOFEM. (GTR-INT-344)*. Ogden, Utah: US Forest Service, Rocky Mountain Research Station.
- Reisen, F., Duran, S. M., Flanniga, M., & Elliott, C. T. (2015). Wildfire smoke and public health risk. *International Journal of Wildland Fire*, 24, 1030–1044. <https://doi.org/10.1071/WF15034>
- Riley, K. L., Grenfell, I. C., & Finney, M. A. (2016). Mapping forest vegetation for the western United States using modified random forests imputation of FIA forest plots. *Ecosphere*, 7, e01472. <https://doi.org/10.1002/ecs2.1472>

- Robinson, A. P., & Froese, R. E. (2004). Model validation using equivalence tests. *Ecological Modelling*, 176, 349–358. <https://doi.org/10.1016/j.ecolmodel.2004.01.013>
- Rollins, M. G. (2009). LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. *International Journal of Wildland Fire*, 18, 235–249. <https://doi.org/10.1071/WF08088>
- Rollins, M. G., Keane, R. E., & Parsons, R. A. (2004). Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. *Ecological Applications*, 14, 75–95. <https://doi.org/10.1890/02-5145>
- Scott, J. H., Thompson, M. P., & Calkin, D. E. (2013). *A wildfire risk assessment framework for land and resource management. (RMRS-GTR-315)*. Ogden, Utah: US Forest Service Rocky Mountain Research Station.
- Shankar, U., Prestemon, J. P., McKenzie, D., Talgo, K., Xiu, A., Omary, M., et al. (2018). Projecting wildfire emissions over the south-eastern United States to mid-century. *International Journal of Wildland Fire*, 27, 313–328.
- Skowronski, N. S., Clark, K. L., Duveneck, M., & Hom, J. (2011). Three-dimensional canopy fuel loading predicted using upward and downward sensing LiDAR systems. *Remote Sensing of Environment*, 115, 703–714. <https://doi.org/10.1016/j.rse.2010.10.012>
- Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., et al. (2013). Carbon stock and density of northern boreal and temperate forests. *Global Ecology and Biogeography*, 23(3), 297–310. <https://doi.org/10.1111/geb.12125>
- Turley, M. C., & Ford, E. D. (2009). Definition and calculation of uncertainty in ecological process models. *Ecological Modelling*, 220, 1968–1983. <https://doi.org/10.1016/j.ecolmodel.2009.04.046>
- Urbanski, S. (2014). Wildland fire emissions, carbon and climate: Emissions factors. *Forest Ecology and Management*, 317, 51–60. <https://doi.org/10.1016/j.foreco.2013.05.045>
- 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 6 February 2019]
- Welsh, A. H., Cunningham, R. B., Donnelly, C. F., & Lindenmayer, D. B. (1996). Modelling the abundance of rare species: Statistical models for counts with extra zeros. *Ecological Modelling*, 88, 297–308. [https://doi.org/10.1016/0304-3800\(95\)00113-1](https://doi.org/10.1016/0304-3800(95)00113-1)
- Wright, C. S., & Eagle, P. C. (2007). Digital Photo Series. Final Report to the Joint Fire Science Program. Boise, ID. [Accessed 6 February 2019 at https://www.firescience.gov/projects/04-4-1-02/project/04-4-1-02_final_report.pdf]