

Modeling Land use Change and Forest Carbon Stock Changes in Temperate Forests in the United States

Lucia Alejandra Fitts (✉ fitts010@umn.edu)

University of Minnesota <https://orcid.org/0000-0003-3645-8710>

Matthew B. Russell

University of Minnesota Department of Forest Resources

Grant M. Domke

USDA Forest Service Northeastern Research Station: USDA Forest Service Northern Research Station

Joseph K. Knight

University of Minnesota Department of Forest Resources

Research

Keywords: Ecosystem services, carbon dynamics, forest loss drivers, forest inventory, remote sensing, USDA Forest Inventory and Analysis (FIA) data

Posted Date: November 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-104361/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

Forests provide the largest terrestrial sink of carbon (C). However, these C stocks are threatened by forest land conversion. Land use change has global impacts and is a critical component when studying C fluxes, but it is not always fully considered in C accounting despite being a major contributor to emissions. An urgent need exists among decision-makers to identify the likelihood of forest conversion to other land uses and factors affecting C loss. To help address this issue, we conducted our research in California, Colorado, Georgia, New York, Texas, and Wisconsin. The objectives were to (1) model the probability of forest conversion and C stock's dynamics using USDA Forest Service Forest Inventory and Analysis (FIA) data and (2) create wall-to-wall maps showing estimates of the risk of areas to convert from forest to non-forest. We used two modeling approaches: a machine learning algorithm (random forest) and generalized mixed-effects models. Explanatory variables for the models included ecological attributes, topography, census data, forest disturbances, and forest conditions. Model predictions and Landsat spectral information were used to produce wall-to-wall probability maps of forest change using Google Earth Engine.

Results:

During the study period (2000-2017), 3.4% of the analyzed FIA plots transitioned from forest to mixed or non-forested conditions. Results indicate that the change in land use from forests is more likely with increasing human population and housing growth rates. Furthermore, non-public forests showed a higher probability of forest change compared to public forests. Areas closer to cities and coastal areas showed a higher risk of transition to non-forests. Out of the six states analyzed, Colorado had the highest risk of conversion and the largest amount of aboveground C lost. Forest disturbances were not a major predictor of land use change.

Conclusions:

Land use change is accelerating globally, causing a large increase in C emissions. Our results will help policy-makers prioritize forest management activities and land use planning by providing a quantitative framework that can enhance forest health and productivity. This work will also inform climate change mitigation strategies by understanding the role that land use change plays in C emissions.

Background

Forests, as part of their ecosystem services, serve as the world's largest terrestrial sink of carbon (C) by storing it in biomass and soil (1–3). This C cycles through the ecosystem via biogeochemical processes causing it to move between different pools, (i.e., aboveground and belowground biomass, dead wood, litter, organic soil, and harvested wood products) or back to the atmosphere depending on the ecosystem's dynamics and disturbances. These processes include photosynthesis, respiration,

decomposition, and natural and anthropogenic disturbances (4). Studies of C stocks in forests are key to informing changes in greenhouse gas emission and removal accounts under climate change scenarios (3) and can be estimated at different scales.

When quantifying C at a large geographic scale, land use and land cover change and disturbance history are essential components to consider (3, 5, 6). In this matter, land use dynamics (2, 6–8) including land use legacies (7, 9), are major factors affecting terrestrial C fluxes. For example, C accumulation in forests of the eastern United States has been credited to historical changes in land use, especially to forest regrowth after agricultural abandonment (9). However, when quantifying C at a stand-level scale, land use change is rarely incorporated (10), creating uncertainties in C accounting.

Carbon emissions due to land use change can be quantified by separating the different C fluxes into its individual components. The Intergovernmental Panel on Climate Change (IPCC) provides guidelines for the estimation of national greenhouse gas (GHG) inventories (11), including C, which are consistent with reporting requirements in the United Nations Framework Convention on Climate Change (UNFCCC). These guidelines specifically include operational models to estimate fluxes (emissions and removals) using a process for each C pool. The fluxes are estimated for each land use category and differentiate the categories remaining the same from those that were converted to another land use.

However, much uncertainty exists when quantifying how much land use and land cover change actually contributes to C flux (2, 12), due, in part, to a lack of confidence in separating these fluxes into individual components (12). In addition, C dioxide emissions or transfers resulting from land use change may be underestimated as some processes (e.g., tree harvesting and land clearing from shifting cultivation) are not considered (9, 12–14) or data is limited. Similarly, there are too few global-scale observational constraints to exclusively estimate anthropogenic land use and land cover C emissions (12). Different methods have been used to estimate changes in C density caused by land use change. Three of the most common approaches include: inventory-based estimates, satellite-based estimates, and process-based vegetation models (14). In the US, estimates of GHG emissions and removals are estimated with data from the Forest Inventory and Analysis (FIA) program, which is in charge of conducting the US national forest inventory (4).

One of the common measures used in land use change research is land cover change, though this does not always accurately reflect actual land use change (15). A crucial difference between these two concepts is that tree cover loss (in case of forest cover) does not always show the activities that actually happen on the ground. Different drivers affect land use and land cover change. On the one hand, global drivers of tree cover loss include deforestation (mostly in Southeast Asia and Latin America), shifting agriculture (Africa and Latin America), forestry (Europe, North America, Russia/China/Southeast Asia, and Australia/Oceania), wildlife (Russia/China/Southeast Asia, Australia/Oceania, and North America), and urbanization (North America) (16). On the other hand, land use change is caused by both human and climate drivers. Decisions on land use are often based on short-term economic factors and are influenced by globalization, technological innovation, and policies at different levels (i.e. local, state or national)

(17). For forest lands, the risk of conversion to other land uses is correlated with environmental, political, social, cultural, and economic factors (10, 17). Key drivers of this conversion include changes in demographic variables (3), urban expansion (18), distance to the nearest road (10), and deforestation for commodity production (16). Therefore, understanding the trends and long-term demographic context for population change could aid land managers and other stakeholders in mitigating the effects of residential development, especially near public lands, and anticipate future human population changes (19).

While global projections on carbon are bleak, the current situation in the US shows a better picture. Land use change at a global level is projected to contribute between 11 and 110 billion metric tons of carbon to the atmosphere by 2050 due to economic, social, and demographic trends (8). For forest land, global trends indicate a loss in the tropics and an increase in Europe, China, and North America (14). Specifically in the US, this trend is due to better forest management practices, reforestation, and an improvement in natural resources management, which have contributed to 11–13% of the global ecosystem carbon removal (8). However, even though these forests are expected to continue sequestering carbon, they would do so at declining rates mainly due to aging forests and land use dynamics (20).

The US's future in C emissions might change according to land use change projections. Some studies show that even though current estimates for the eastern US (2001 to 2012) indicate that forest land use has changed (positive or negative) less than 5%, large changes in land use are likely in the coming decades in a business-as-usual scenario (3). For this study, recent trends indicated increasing forest areas in the southern Plains and Great Lakes' states and losses in forest areas in the central and south-central states. Additional areas with high probability of conversion to non-forest include the Great Plains, especially in poorly stocked areas and/or sites with small diameter trees (10). Other studies estimate as high as 36% of the land area in the conterminous United States to change in land use between 2001 and 2051 in a business-as-usual scenario (21). According to this study, urban and forest land use in the US are predicted to increase by 79% and 7%, respectively. On the other hand, cropland and pasture land use are expected to decline by 16% and 13%, respectively (21).

Overall, the study of land use change is critical in forest C dynamics and better land use planning is needed to secure ecosystem services provided by forests. Even though some C stocks are increasing due to forest regrowth, especially from agricultural abandonment (22, 23), it is critical that we address the issue of forest conversion due to its significant contribution on the C budget. For example, in the US, forest lands, harvested wood products, and urban trees together offset more than 11% of the total annual GHG emissions (24). Here, C uptake estimates for forests remaining forests were - 564.5 MMT CO₂ equivalent (eq.) in 2018. On the other hand, C emission estimates for forests which transitioned to other land uses during the same year were 127.4 MMT CO₂-Eq. (24). To help address these issues, our research focuses on modeling the relationships among disturbances, land use change, and aboveground C stocks in six US states over an 18-year period. The research objectives are to (1) model the probability of forest conversion and C stock's dynamics using USDA Forest Service Forest Inventory and Analysis (FIA) data and (2) create wall-to-wall maps showing estimates of the risk of areas to convert from forest to non-

forest. This research will allow us to better understand the impacts of land use change on the forest C cycle and be able to more effectively determine priority areas for management and land use planning.

Methods

Study area

This research includes six US states: California, Colorado, Georgia, New York, Texas and Wisconsin (Fig. 1). These six states were chosen because they are spread throughout the country and represent each of the US FIA work units (25). This diversity in locations allowed us to represent different forest types, disturbance history, population dynamics and drivers that could influence our study variables. In addition, there were at least two plot remeasurements for all of those states, which allows for a temporal analysis of change. The study area was limited to areas with forest land cover at the start of the study.

Forests in the US cover more than one-third of its landscape and comprise more than 28 billion cubic meters of wood volume. Even though overall forest area has remained stable, there have been changes at regional and local scales that are not reflected by aggregated numbers. In addition, the road network in the country has expanded, making forest lands more accessible to people (26). This easier access to forests would likely encourage land use changes. Major natural forest disturbances in the area include wildfire, insects, and disease (26). These and other disturbances, as well as changing forest conditions, cause a change in tree species composition and distribution. Climate in the study region varies widely (Table 1).

Table 1
Climate for each state's capital city analyzed in this study (27).

State	Average maximum temperature (°C)	Average minimum temperature (°C)	Maximum temperature (°C)	Minimum temperature (°C)	Total precipitation (mm)
California	23.1	9.1	33.3	3.3	470
Colorado	18.3	2.5	32.2	-7.8	360
Georgia	22.2	11.8	32.2	1.1	1263
New York	14.3	3.7	27.8	-9.4	999
Texas	26.6	15.0	36.1	5.6	870
Wisconsin	13.2	2.7	27.8	-11.7	876

Data Collection

Ground data used in this study were collected and organized into a database by the USDA Forest Inventory and Analysis (FIA) program from 2000 to 2017. This database includes different phases and

tables depending on the scale and level of detail of the information contained in each. For this research, we worked with ground-sampled FIA plots that were remeasured approximately every five years for eastern states and 10 years for western states, which allowed for a spatiotemporal analysis. These plots covered a 0.4 hectare sample area and were divided into four fixed-radius (7.32 m) subplots. Within each subplot, a microplot (2.07 m radius) was nested. Here, trees with a diameter at breast height (d.b.h.) less than 12.7 cm were measured. Forest and non-forest conditions were assigned by the field crew in each remeasurement, which allowed us to track any land use changes that occurred. To obtain human population and housing information, we used the United States' 2010 National Census (19). For obtaining spectral information, we used a mosaic of Landsat 8 surface reflectance imagery (30 m spatial resolution) from the period 2018-05-01 to 2018-09-30 (which corresponded to the summer months) obtained from the Google Earth Engine platform (LANDSAT/LC08/C01/T1_SR). The initial image collection consisted of 4093 Landsat images that covered the study region during the summer months.

Data analysis

Information at the subplot and microplot levels was gathered for the different time periods available (initial time t_i and final time t_f) and scaled-up using the FIA expansion factors to a plot level (minimum modeling unit). The initial data set was divided into training (0.7) and test (0.3) data sets to fit and validate the models. Our goal was to build one model per variable of interest (response variable). Therefore, model comparisons were performed to select the model which yielded the optimal fit statistics. The first attribute of interest was the probability of forested areas being converted to mixed or non-forested areas. For the purposes of this study, we will refer to forested plots which transitioned to mixed (e.g., one forest plus one non-forest condition) or non-forest conditions as plots that changed in land use. The second variable was the C stock changes in the total aboveground live biomass, defined as the sum of dry biomass located in the merchantable bole, top and limbs, stump, woodland tree species and saplings.

Model building

To explore the performance of different modeling approaches, we compared two main methods: a statistical model and a machine learning algorithm. For both models, we compared the performance of these two main approaches to select the best performing model for each variable of interest. For the machine learning algorithms, we opted to use a random forest algorithm due to its capability of working with a large dataset and variables of interest, as well as its capability of working well for both categorical (through classification procedures) and quantitative (through regression procedures) variables (28). The random forest algorithm was used for both models: change in C stocks and land use change. For the statistical models, we used mixed effects modeling because we worked with ecological data that had nested data, as well as temporal and spatial correlation structures. To select which type of mixed effects models to use, we looked at the characteristics of our variables of interest, especially the response variable. For the land use change model, the response variable had a binary behavior (0: no change and 1: change). Therefore, we chose a logistic model, which would give us the probability (or odds) of the

response variable. In addition, for the C model, the response variable was quantitative and had a linear relationship with the predictor variables. Therefore, we selected a linear regression model that accounted for correlated data, a linear mixed effects model with a random intercept.

Random forest model

For both models (land use change and C), our first approach was to use the random forest machine learning algorithm. This algorithm consists of building multiple decision trees (collectively called a forest), each of which will work with a random sample of the data. To avoid bias due to outliers, the random forest algorithm will take the average prediction from all the decision trees (i.e., the mean for regression and the mode for classification) by adopting a bagging approach (29–31).

We used the `randomForest` function in R (32) to run 500 trees and create a model with all the explanatory variables initially selected (including ecological, topographical and census variables). The advantage of this algorithm is that it can work with an extensive amount of categorical and quantitative predictor variables. For the land use change model, as the response variable was categorical, the algorithm predicted a classification. For the C model, a quantitative response was estimated. We then identified the importance of each variable in the model. For the land use change model, a confusion matrix was created to determine the ability of the model to capture the conversion to and from the different categories.

Generalized linear mixed effects models

The generalized linear mixed effects models unify several different regression models (i.e. general linear models, logistic regression, Poisson regression) while accounting for spatial and temporal correlations and nested data. This family of models has two main components: fixed effects (which includes the variables we are interested in) and random effects (which acknowledge the effect of the variables included here with a variation that is normally distributed with a certain variance) (33, 34).

For both the LUC and C models, we divided our data into training (0.7) and test (0.3). The training data were used to create a generalized mixed-effects model for each of the response variables. Explanatory variables for both models were divided into the following categories: 1) forest attributes (ecological variables), 2) plot attributes, 3) topography, 4) census data, 5) forest disturbances, and 6) number of conditions present in a plot (e.g., number of forest and non-forest types differentiated by reserved status, owner group, forest type, and stand density). We regrouped the ecological classification codes into broader ecoregions due to the large number of categories, which would have complicated the model.

Variable selection techniques were applied in two stages. The first stage included dividing the variables into the categories defined above. Within each category, Pearson correlation coefficients were calculated to quantify relationships between quantitative variables. Variables with a coefficient of 0.5 or higher were marked to later verify that no correlated pair were kept in the final model. Categorical variables were visually assessed through scatterplot matrices. No variables were discarded in this step. In addition, for the logistic model we used the information value (IV) indicator to see how strong of a predictor each

variable was and in which order to include the variables in the model (Table 4). Variables were added to the model one at a time in descending order of their IV level of importance. For the regression model, variables were incorporated into the model according to the variable importance ranking obtained from the random forest model (Additional file 1). For both models, the Akaike information criterion (AIC) was used for model comparison. If an added variable increased the AIC value, it was removed. After this first variable selection stage, variables kept in each category were merged back to proceed with the variable selection.

We compared outputs of three variable selection approaches in R software using the *car* (35) and *MASS* (36) packages: forward selection, backward selection and stepwise selection. Finally, we used the variance inflation factor (VIF) to verify there was no multicollinearity ($VIF > 10$) present in the model. A logistic regression model with the remaining variables was created for the land use change model using the *glmPQL* function in R with the *cloglog* link (36). In contrast, a linear mixed-effects model was implemented for the C model using the *glmer* function (37). For both models, different combinations of the retained variables plus state and forest types or ecoregion (independently and nested within each other) in the fixed or random section of the model were validated with the test dataset. This was done to account for correlated data points in our spatiotemporal analysis. A complete workflow of the process can be observed in Fig. 2.

Model comparison

For both generalized mixed-effects models (logistic and linear models), a comparison between these techniques and the random forest approach was done to identify the best performing model. The criteria used to select the model was to look at the percentage of omission and commission errors, overall accuracy, precision, and recall values. While we used the AIC values to compare models at the initial stage (to filter out preliminary models, Fig. 2), once we incorporated random effects in the models, AIC values were no longer estimated. Since we used the *glmPQL* function in R to compute the models, no quasi-likelihood can be estimated for these mixed-effects models (38). In the case of the C model, the root mean square error (RMSE) was used for model selection.

Creating predicted wall-to-wall likelihood of change

A wall-to-wall map for the study area using the FIA data and the logistic model equation (See model building section) was created to predict the likelihood of conversion of forested areas for every pixel. We only used the explanatory variables from the logistic model that had information available at a wall-to-wall scale. These variables were obtained through Google Earth Engine (GEE) and ArcMap 10.8 using original FIA coordinates, mosaic Landsat 8 images, and elevation data. The variables used were: Normalized difference vegetation index -NDVI- (obtained from the Landsat 8 mosaic) as an estimate of basal area, ownership code (39), aspect (40), remeasurement period (Mean value per state from the FIA plot data), and census data (19). All layers were converted to a raster format and transformed to the same projection (WGS84).

The NDVI variable was calculated in the GEE platform. Bands 5 and 4 from the imported Landsat 8 Surface Reflectance images for the summer months of 2018 for each state were used to generate a maximum NDVI for each pixel. This was done to obtain a mosaic of pixels displaying the maximum NDVI value for the period selected which would avoid clouds and ensure a value where hardwood trees had leaves. To extract NDVI and aspect values for each pixel, we used true FIA plot coordinates. Means for plot coordinates were calculated to account for variability and errors in each coordinate measurement. A buffer of 180 feet was used around the plot center to replicate the extent of a plot and calculate mean values for each pixel through zonal statistics.

With this wall-to-wall information, a multiple linear regression model was fit in R (with the predicted probability of change previously calculated as a response variable, see “Model building” section), obtaining an equation for the linear model. This equation was used with the raster calculator tool in the ArcGIS environment. The maps generated here were clipped with the National Land Cover Database (NLCD) to display only forested areas. This new raster layer displayed the probability of conversion from forest to another land use (See Fig. 3 for a visual workflow).

For interpretation of the results shown in the maps, we overlapped ownership, protected areas, digital elevation models (DEM), and major cities' shapefiles to our maps.

Results

The initial exploration of the FIA data for the six states revealed some important foundational knowledge. From 2000 until 2017, California, Colorado, New York, and Wisconsin had an increase in mean aboveground live biomass, while Georgia and Texas had a mean reduction in live biomass. On the other hand, mean biomass for standing dead trees increased in California, Georgia, New York, and Wisconsin (Table 2). We also observed that 3.4% of forested plots transitioned from forest to mixed or non-forested conditions and 0.7% of the plots transitioned from forest or mixed conditions to non-forest. California had the highest percentage of plots that changed in land use, followed by Texas (Table 2 and Fig. 4).

Table 2
Estimates for aboveground biomass and change from forest land use, FIA data

State	Change in aboveground live biomass (Mg/ha)	Change in aboveground dead biomass (Mg/ha)	Number of plots evaluated	Number of plots that transitioned from forest to mix or non-forest condition	Percentage of plots that transitioned per state
California	0.36 ± 2.71	0.72 ± 5.14	954	79	8.28
Colorado	0.72 ± 2.44	-0.98 ± 3.20	646	7	1.08
Georgia	-0.02 ± 1.18	0.54 ± 7.88	3139	100	3.19
New York	0.05 ± 1.13	0.78 ± 4.19	1887	47	2.49
Texas	-0.05 ± 1.16	-0.26 ± 7.65	2067	98	4.74
Wisconsin	0.06 ± 0.98	0.49 ± 3.93	5632	149	2.65
All states	0.38 ± 5.73	0.08 ± 1.35	14325	480	3.35

Legend: Mean estimates (± standard deviation) and plot information for the variables of interest across six states in the US, 2000–2017

Figure 4: Percent of FIA plots across six US states that have transitioned to and from forested conditions in the study area between 2000–2017.

Land use change model building

Random forest classification algorithm

Forested plots that remained forested were well predicted by the random forest model (classification error was around 0.1%). However, the model failed to predict the plots that changed in land use (classification error over 99%) (Table 3). The overall out-of-bag (OOB) estimate of error rate was 3.43%.

Table 3
Confusion matrix predicting forest land use change for the random forest

Class	No change	Change	Classification error
No change	9712	11	0.0011
Change	334	3	0.9911

Legend: Random forest model was done with 500 trees and 5 variables tried at each split in six US states between 2000–2017

The five most important variables for the random forest classification model were live basal area, live trees per hectare, forest type, elevation, and dead basal area (See Additional file 1 for a complete ranking

of variable importance).

Logistic regression model

Out of the 32 initial variables (Table 4, see Additional file 2 for metadata), nine were retained after the selection process for fixed effects: basal area live, basal area dead, remeasurement period, ownership code, percentage of protected areas per county, natural increment per county, number of forested conditions in a plot, number of non-forested conditions in a plot, and aspect. For random-effects, forest type was a stronger predictor than ecoregion (IV of 0.52 compared to 0.35 for ecoregion) (Table 4).

Table 4
Variables used in the land use change model displaying their strength as predictor variables

Strong (IV < 0.02)	Medium (0.02 ≤ IV < 0.1)	Weak (0.1 ≤ IV < 0.3)	Not useful predictor (IV ≥ 0.3)
Ecological classification code (0.65)	Basal area per hectare live (0.3)	Natural increment (0.1)	Aspect (0.06) Disturbance: Simple/Compound (0.01)
Forest type (0.52)	Physiological classification code (0.2)	Ownership code (0.09)	Change in population (0.06)
Forest type group (0.39)	Basal area per hectare dead (0.2)	Percentiles of protected land (0.09)	Ownership group code (0.06)
Ecoregion (0.35)	Trees per hectare dead (0.2)	% of county land publicly administered (0.08)	Elevation (0.05)
Remeasurement period (0.3)	State (0.18)	House density (0.08)	Number of forested conditions (0.04)
	Trees per hectare live (0.13)	Distance to the nearest road (0.07)	Main disturbance (0.04)
	%Protected areas (0.1)	Number of non-forested conditions (0.07)	Net migration (0.03)
		Percentiles of public land (0.07)	Percent change in housing units (0.03)
		Slope (0.06)	Water code (0.02)
Legend: Information value shown in brackets, numbers rounded to the nearest two decimal places.			

Results from the validation of the logistic model show that class 0 (no forest land use change) had a very low omission and commission error (between [0.05 – 0.02] and [3.26–3.33], respectively). However, for class 1 (change from forest to other land use), the omission error was high (close to 100% in most cases). This might have been due to the model having difficulty capturing class 1 change because there was a higher proportion of FIA plots that did not change in forest land use between t_i and t_f compared to the number of plots that did change. Overall accuracy was over 96%, which should be interpreted with caution as it might be driven by the greater proportion of correctly classified 0's. Comparing all these indicator variables, the logistic regression was selected with forest type nested within state as random effects. Coefficients and equations for the final model can be observed in Additional file 2 and predictions in Fig. 5.

Table 5

Model comparison between random forest and the final candidates for the logistic regression models

Model	Accuracy	% Omission	% Commission	Class	Precision	Recall	
<i>Random forest</i>	96.589	0.05	3.37	0	0	0	
		100	100	1	100	1	
<i>Logistic model</i>	<i>State as a fixed effect</i>	96.68	0.02	3.3	0	71.43	1.48
			98.52	28.57	1	28.57	1
	<i>Forest type as a fixed effect</i>	96.7	0.03	3.27	0	72.73	2.37
			97.63	27.27	1	27.27	1
	<i>State as a random effect</i>	96.68	0.02	3.3	0	71.43	1.48
			98.52	28.57	1	28.57	1
<i>Forest type as a random effect</i>	96.69	0.03	3.28	0	70	2.08	
		97.92	30	1	30	1	
Forest type nested within State as a random effect	96.71	0.03	3.26	0	75	2.67	
		97.33	25	1	25	1	

Legend: Bold text indicates the selected model. Category 0 indicates no change from forest land use, while class 1 indicates a change from forest land use

Figure 5: Predictions from the logistic regression model. Response variable is the probability of forest land use change to mixed or non-forest conditions. The four horizontal panels show different land ownership classes. Figure 5a. The probability of land use conversion for the six states analyzed. Figure 5b. The effects of population natural increment. Figure 5c. The effects of the percentage of protected areas in a county.

Colorado had the highest probability of forest conversion compared to the other five states followed by California. Areas with high population natural increment and the lowest percentage of protected areas had a higher predicted probability of conversion. In addition, non-public areas had a higher probability of conversion to non-forested areas (Fig. 5).

Aboveground live carbon model building

Random forest regression algorithm

The five most important variables for the random forest regression model were live basal area, live trees per hectare, elevation, dead basal area, and dead trees per hectare (See Additional file 1 for a complete ranking of variable importance).

Mixed-effects linear model

Out of the 36 initial variables (Additional file 3), after the selection process for fixed effects, nine variables were kept: basal area live, basal area dead, trees per hectare live, distance to roads, disturbances (simple/compound), percentage of protected areas per county, number of forested conditions in a plot, condition at t_f , and change from forest to other land use. Different combinations of linear models with these variables plus state and forest types or ecoregion in the fixed or random section of the model were validated with the test dataset. Forest type was a stronger predictor than ecoregion when comparing AIC; hence, forest type nested within state was used as the random effect.

Coefficients and the equation for the final model can be observed in Additional file 3. The mixed-effects model had a lower root mean square error (RMSE) compared to the random forest model (12.06 compared to 13.56 MG/ha); hence, the mixed-effects model was the one selected to predict change in aboveground live C stocks.

Predictions from best performing model

Colorado and California had a more negative change in aboveground live C compared to the other states (Fig. 6). Areas with the lowest percentage of protected areas had a more negative change in C. In addition, disturbances had an effect on the aboveground C stocks. We observed that plots with more than one disturbance in between remeasurements (compound disturbance) had a more negative change in C than plots being affected by one (simple) disturbance.

Creating predicted wall-to-wall likelihood of change

Public areas (e.g., forests managed by federal and state agencies) have a lower probability of conversion to non-forest (Fig. 7). Protected areas within the six states show lower probability of conversion to non-forested areas. Second, we see that areas closer to big cities and coastal areas tend to have a mid to high probability of conversion. Topography plays an important role, in that higher elevation areas where accessibility is difficult are represented by areas that have a low probability of conversion to non-forested

areas (e.g., mountain ranges). Fourth, in most of the states, areas closer to rivers tend to have greater risk to convert to non-forested areas (moderately high to high probability).

Discussion

Our results show that land use change affects forested areas, especially due to growing populations and urban development. Urbanization is a main driver of the loss of forests in the US (16, 18), which brings agricultural expansion and concerns for food security. Other studies specifically report demographic variables (3) and proximity of roads (10) as important urbanization variables that affect forests. Given the increase in housing relative to population growth in the US (19), demographic variables are critical in estimating forest conversion and C emissions. Through the application of the C model, we observed that forest conversion affects C fluxes, especially at large scales. This relationship between land use change and C fluxes has been reported in other studies (2, 3, 5–7). Therefore, it is critical to incorporate estimates of land use change in C studies and C in land use change studies.

Although we observed that less than 4% of the FIA plots transitioned from forest to other land uses, prior studies (3) have shown that large changes are likely to occur in the coming decades in a business-as-usual scenario in eastern US forests. Similarly, around 36% of the forested area in the conterminous US area is estimated to change in land use (21). Therefore, it is key to identify critical regions that have a higher probability of conversion, despite the actual probability being low relative to other risks.

An important result of this research is that demographic variables (such as the population natural increment), ownership of the land, level of protection, and live and dead biomass (represented through basal area and trees per hectare) represent important factors driving forest conversion. With respect to ownership of the land, motivation to sell land or change its use might play an important role in private lands (41, 42), as well as short-term economic factors and technological innovation (17). These reasons could explain the differences in probability of conversion observed in this study among ownership groups.

Disturbances were not a major factor when modeling the probability of forest conversion for all the states together. However, looking at each state independently, California and Colorado had a stronger influence of the variable forest disturbances which may be due to widespread insects, diseases, and fire in those states (43). Even though the disturbance variables were not kept in the model, their influence in these states might be captured by the state coefficients (used as random effects). These disturbances might explain the difference observed in Fig. 5a, in which both states show a higher probability of conversion compared to other states. On the other hand, when estimating aboveground C stock changes, the presence of disturbances decreased C fluxes whether those disturbances were simple or compound.

From the wall-to-wall maps, we observe that California's natural protected areas such as federally-managed forest lands (e.g. Six Rivers, Shasta Trinity, Sierra National Forests, and Yosemite National Park) display a low probability of conversion (mean: $0.0445 \pm$ standard deviation: 0.0135). A surprising finding was that the Sequoia National Forest (0.0575 ± 0.0055) has a moderately high probability of

conversion to non-forest. This area has had a very high natural increment in its population (19). In addition, coastal areas exhibit a moderately high probability of conversion, while mountain areas exhibit a low probability.

Compared to other states, areas surrounding rivers in Colorado have a low probability of conversion. This area is characterized by an abrupt topography, similar to mountain ranges, which also have a lower probability of conversion. Colorado's natural protected areas such as Rio Grande National Forest, which has very few small towns in its surroundings, present a low probability of conversion (0.0299 ± 0.0095). However, the Colorado Springs area has a moderately high probability of conversion to non-forest (0.0598 ± 0.0083). This area has a high natural increment in its population (19), which would explain this probability. Costilla and Las Animas counties display a high probability of conversion (0.0726 ± 0.0067 and 0.0635 ± 0.0008 respectively). There is a low percentage of protected areas in these counties as well as more private than public forest lands. Concerning modeled C stocks, both California and Colorado show separation from the rest of the states (following the same trend as the logistic model of forest conversion) and showed a more negative net change in aboveground C stocks.

In Georgia, areas around the Appalachian Mountains represented a low probability of conversion, as well as protected areas such as the Okefenokee National Wildlife Refuge (0.0185 ± 0.0071). Areas surrounding the city of Atlanta have a moderate to high probability of conversion, as well as areas surrounding rivers. The state is characterized by a low percentage of protected areas per county and a high natural increment of its population in most counties.

Forest ownership in New York is similarly distributed between private and public areas (39), making this an important variable that drives forest conversion in the state. Protected areas such as the West Canada Lake Wilderness and the High Peaks Wilderness present a low probability of conversion (0.00 ± 0.0083 and 0.0106 ± 0.0093 respectively). In addition, areas like the Big Indian Wilderness and Allegany State Park have a low probability of conversion as well (0.0250 ± 0.0093 and 0.0218 ± 0.0119), similar to areas federally owned. In general, there is a low percentage of protected areas in the state.

Coastal areas in Texas and areas close to major cities like Austin, Dallas, Houston and San Antonio have a higher risk of conversion. The Big Thicket National Preserve Area (0.0281 ± 0.0019) and the Sabine National Forest (0.00 ± 0.0050) area present a low probability of conversion. The eastern border of the state (where pine forests are located) is predicted to remain as forests. Areas that surround big lakes and cities show a moderately high probability of conversion to non-forest.

In Wisconsin, an important result that stands out in Fig. 7 is the dark green and well-defined area that makes up the Menominee reservation. This area is predicted to remain as forest, as represented by its low probability of conversion (0.00 ± 0.0054). Several studies and publications have highlighted the hard forest management labor that this reservation has been doing (44–46) which is reflected in the model presented in this research. Moreover, protected areas such as the Black River State Forest (0.0090 ± 0.0060), the Apostle Islands (0.00 ± 0.0055), and the Chequamegon National Forest (0.0018 ± 0.0080) display a low probability of conversion. Similar to other states, areas close to the Great Lakes (e.g., Lake

Michigan), rivers (e.g., the Mississippi and the Wisconsin Rivers), and major cities (e.g., Milwaukee) have a higher risk of conversion.

Conclusions

Land use change and resulting contributions to the global C budget are accelerating. A better understanding of the drivers of land use change is needed to reduce the loss of forested areas and C emissions, as well as to increase the resistance and resilience of forests. From our study, we saw that the main drivers of forest conversion in FIA plots were the live and dead biomass present, the amount of protected land, the natural increment of the population, and ownership of the land. An important finding was that disturbances were not a strong predictor of forest conversion, except in California and Colorado where disturbance effects might be captured through the state coefficients. Moreover, resiliency of aboveground C stocks mainly depended on the live and dead biomass present in the forest and its structure, the elevation of the area, the disturbance history, and the area under protection.

Our results will help policy-makers prioritize forest management activities and land use planning by providing tools and spatial information that will help them create forest management strategies that enhance forest health and productivity. Those strategies could include promoting extension programs that reinforce the value of forest, provide economic incentives towards forest management, and implement new technology for monitoring and harvesting resources. These programs should be specially targeted to landowners.

Moreover, to mitigate climate effects on forests, this research highlights priority areas where more intervention is needed to keep these areas as forest despite the changing conditions and increase in disturbances. This research will directly assist six US states (CA, CO, GA, NY, TX, and WI) and if expanded, the benefits can be applied to the entire United States or even worldwide because of the scale at which land use change problems occur.

Overall, this study confirms that land use change is associated with C fluxes and that the zoning of urban and agricultural areas is key to avoid an increase of C emissions that contribute to climate change. Food security is a big concern, which creates an inevitable need to dedicate more areas for residential purposes. However, through more focused forest management that increases the resistance and resilience of ecosystems and prioritizes critical areas, forest health, and productivity, we can create strategies to balance population and ecosystem needs.

Declarations

Availability of data and materials: The data that support the findings of this study are openly available in the USDA Forest Service FIA Datamart public data repository at: <https://apps.fs.usda.gov/fia/datamart/>

Competing interests: The authors do not have any competing interests to declare.

Funding: This work was supported by the US Department of Agriculture - Forest Service, Northern Research Station (project 14-JV-11242305-012), the Minnesota Agricultural Experiment Station (project MIN-42-101), and University of Minnesota Department of Forest Resources fellowship program.

Authors' contributions: LAF, MBR, and GMD conceived and designed the study. LAF conducted the analyses and wrote a first draft of the work. MBR, GMD, and JKK provided input on the analysis and interpretation of results. All authors read and approved the final manuscript.

Corresponding author: Correspondence to Lucia Fitts at fitts010@umn.edu

Authors' information: LAF: ORCID iD: [0000-0003-3645-8710](https://orcid.org/0000-0003-3645-8710), fitts010@umn.edu; MBR: ORCID iD: 0000-0002-7044-9650, russellm@umn.edu; GMD: grant.m.domke@usda.gov; JKK: jknight@umn.edu

Acknowledgements: We would like to express our gratitude to The USDA Forest Service Northern Research Station, especially to Brian Walters for the help with data access and processing, to the University of Minnesota Department of Forest Resources, and to The Student Writing Support at the University of Minnesota.

References

1. Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, et al. A Large and Persistent Carbon Sink in the World's Forests. *Science* (80-). 2011;317(july):4.
2. Sleeter BM, Liu J, Daniel C, Rayfield B, Sherba J, Hawbaker TJ, et al. Effects of contemporary land-use and land-cover change on the carbon balance of terrestrial ecosystems in the United States. *Environ Res Lett*. 2018;13(4).
3. Woodall CW, Walters BF, Coulston JW, D'amato AW, Domke GM, Russell MB, et al. Monitoring Network Confirms Land Use Change is a Substantial Component of the Forest Carbon Sink in the eastern United States The Growing Question of Land Use and the Forest Carbon Sink in the US. *Nat Publ Gr [Internet]*. 2015 [cited 2018 Dec 5]; Available from: www.nature.com/scientificreports.
4. Domke GM, Walters BF, Nowak DJ, Smith J, Ogle SM, Coulston JW. Greenhouse gas emissions and removals from forest land and urban trees in the United States, 1990–2017. *Resour Updat FS-178 [Internet]*. 2019;2018:2017–20. Available from: <https://www.nrs.fs.fed.us/pubs/57919>.
5. Albani M, Medvigy D, Hurtt GC, Moorcroft PR. The contributions of land-use change, CO2 fertilization, and climate variability to the Eastern US carbon sink. *Glob Chang Biol*. 2006;12(12):2370–90.
6. Woodall CW, Walters BF, Russell MB, Coulston JW, Domke GM, D'Amato AW, et al. A Tale of Two Forest Carbon Assessments in the Eastern United States: Forest Use Versus Cover as a Metric of Change. *Ecosystems*. 2016;19(8):1401–17.
7. Pacala SW, Hurtt GC, Moorcroft PR, Caspersen JP. Carbon Storage in the U. S. Caused by Land Use Change. *Present Futur Model Glob Environ Chang Towar Integr Model [Internet]*. 2001;1(1):145–72. Available from: <https://www.terrapub.co.jp/e-library/toyota/pdf/145.pdf>.

8. Cavallaro N, Shrestha G, Birdsey R, Mayes MA, Najjar RG, Reed SC, et al. Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report. Usgcrp [Internet]. 2018;878 pp. Available from: <https://doi.org/10.7930/SOCCR2.2018>.
9. Caspersen JP, Pacala SW, Jenkins JC, Hurtt GC, Moorcroft PR, Birdsey RA. Contributions of land-use history to carbon accumulation in U.S. Forests. *Science* (80-). 2000;290(5494):1148–51.
10. Puhlick J, Woodall C, Weiskittel A. Implications of land-use change on forest carbon stocks in the eastern United States. *Environ Res Lett*. 2017;12(2).
11. IPCC. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Calvo Buendia, E., Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize, S., Osako, A., Pyrozhenko, Y., Shermanau, P. and Federici S, editor. Switzerland: IPCC; 2019.
12. Arneeth A, Sitch S, Pongratz J, Stocker BD, Ciais P, Poulter B, et al. Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. *Nat Geosci*. 2017;10(2):79–84.
13. Yue C, Ciais P, Houghton RA, Nassikas AA. Contribution of land use to the interannual variability of the land carbon cycle. *Nat Commun* [Internet]. 2020;11(1):1–11. Available from: <http://dx.doi.org/10.1038/s41467-020-16953-8>.
14. Houghton RA, House JI, Pongratz J, Van Der Werf GR, Defries RS, Hansen MC, et al. Carbon emissions from land use and land-cover change. *Biogeosciences*. 2012;9(12):5125–42.
15. Coulston JW, Reams GA, Wear DN, Brewer CK. An analysis of forest land use, forest land cover and change at policy-relevant scales. *Forestry*. 2014;87(2):267–76.
16. Curtis PG, Slay CM, Harris NL, Tyukavina A, Hansen MC. Classifying drivers of global forest loss. *Science*. 2018;361(6407):1108–11.
17. Impacts USGCRP, Risks. and Adaptation in the United States: Fourth National Climate Assessment [Internet]. Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock and BCS, editors], editor. Vol. II. Washington DC: U.S. Government Publishing Office; 2018. 1–470 p. Available from: https://nca2018.globalchange.gov/downloads/NCA2018_FullReport.pdf.
18. Shifley SR, Keith Moser W, Nowak DJ, Miles PD, Butler BJ, Aguilar FX, et al. Five anthropogenic factors that will radically alter forest conditions and management needs in the Northern United States. *For Sci*. 2014;60(5):914–25.
19. Mockrin MH, Stewart SI, Matonis MS, Johnson KM, Hammer RB, Radeloff VC. Sprawling and diverse: The changing U.S. population and implications for public lands in the 21st Century. *J Environ Manage* [Internet]. 2018;215:153–65. Available from: <https://doi.org/10.1016/j.jenvman.2018.03.053>.
20. Wear DN, Coulston JW. From sink to source: Regional variation in U.S. forest carbon futures. *Sci Rep*. 2015;5:1–11.
21. Radeloff VC, Nelson E, Plantinga AJ, Lewis DJ, Helmers D, Lawler JJ, et al. Economic-based projections of future land use in the conterminous United States under alternative policy scenarios. *Ecol Appl* [Internet]. 2012 [cited 2019 Mar 31];22(3):1036–49. Available from: <http://www.nrcs.usda.gov/technical/NRI/>.

22. Caspersen JP, Pacala SW, Jenkins JC, Hurtt GC, Moorcroft PR, Birdsey RA. Contributions of land-use history to carbon accumulation in U.S. Forests. *Science* (80-) [Internet]. 2000 Nov 10 [cited 2018 Feb 12];290(5494):1148–51. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.290.5494.1148>.
23. Shevliakova E, Pacala SW, Malyshev S, Hurtt GC, Milly PCD, Caspersen JP, et al. Carbon cycling under 300 years of land use change: importance of the secondary vegetation sink. *Global Biogeochem Cycles*. 2009;23(2):1–16.
24. EPA. Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990–2017. *J Chem Inf Model*. 2019;Vol. 53:675.
25. Burrill E, Wilson, Am, Turner J, et al. FIA Database Description and User Guide for Phase 2 (version: 8.0) The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2 (version 8.0). 2018.
26. Oswalt SN, Smith WB, Miles PD, Pugh SA. Forest resources of the United States, 2017 [Internet]. General Technical Report - US Department of Agriculture, Forest Service. Forest Service; 2019. Available from: https://www.fs.fed.us/research/publications/gtr/gtr_wo97.pdf.
27. Climate United States. - Monthly Averages [Internet]. [cited 2020 Nov 2]. Available from: <https://www.usclimatedata.com/>.
28. Breiman L. Random forests. *Mach Learn*. 2001;28.
29. Nwanganga F, Chapple M. Practical Machine Learning in R. *Practical Machine Learning in R*. 2020.
30. Ayyadevara VK. Pro Machine Learning Algorithms. Suresh CJ, Moodie M, Modi D, editors. *Statistical Machine Learning*. Springer; 2018. 379 p.
31. Pavlov YL. Random forests. The Netherlands: Ridderprint bv; 2000. 122 p.
32. Wiener AL and M. Classification and Regression by randomForest. *R News* [Internet]. 2002;2(3):18–22. Available from: <https://cran.r-project.org/doc/Rnews/>.
33. Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM *Mixed Effects Models and Extensions in Ecology with R*. Springer Science & Business Media. 2009. 579 p.
34. Kéry M, Schaub M. Bayesian population analysis using WinBUGS: a hierarchical perspective. 2011.
35. Fox J, Weisberg S. An {R} Companion to Applied Regression [Internet]. Second. Sage; 2011. Available from: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
36. Venables WN, Ripley BD. *Modern Applied Statistics with S* [Internet]. Fourth. New York: Springer; 2002. Available from: <http://www.stats.ox.ac.uk/pub/MASS4>.
37. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using {lme4}. *J Stat Softw*. 2015;67(1):1–48.
38. Bolker B. Dealing with quasi- models in R. 2020;(2):5.
39. Hewes JH, Butler BJ, Liknes GC. Forest ownership in the conterminous United States circa 2014: distribution of seven ownership types - geospatial dataset [Internet]. Forest Service Research Data Archive; 2020. Available from: <https://doi.org/10.2737/RDS-2017-0007>.

40. U.S. Geological Survey. 1/3rd arc-second Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection [Internet]. 2017. Available from: <https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5>.
41. Markowski-Lindsay M, Catanzaro P, Milman A, Kittredge D. Understanding Family Forest Land Future Ownership and Use: Exploring Conservation Bequest Motivations. *Small-scale For.* 2016;15(2):241–56.
42. Bengston DN, Service USF, Butler BJ, Asah ST. Values and Motivations of Private Forest Owners in the United States: A Framework Based On Open-Ended Responses in the National Woodland Owner Survey. *Environ Values.* 2008;(May 2014):60–6.
43. Potter KM, Conkling BL. Forest health monitoring: National status, trends, and analysis 2019. 2020; (June):213.
44. Dockry MJ, Hall K, Van Lopik W, Caldwell CM. Sustainable development education, practice, and research: an indigenous model of sustainable development at the College of Menominee Nation, Keshena, WI, USA. *Sustain Sci.* 2016;11(1):127–38.
45. Trosper RL. Indigenous influence on forest management on the Menominee Indian Reservation. *For Ecol Manage.* 2007;249(1–2):134–9.
46. Mausel DL, Waupochick A, Pecore M Menominee forestry: Past, present, future. 2017;90(September):109–24.

Figures

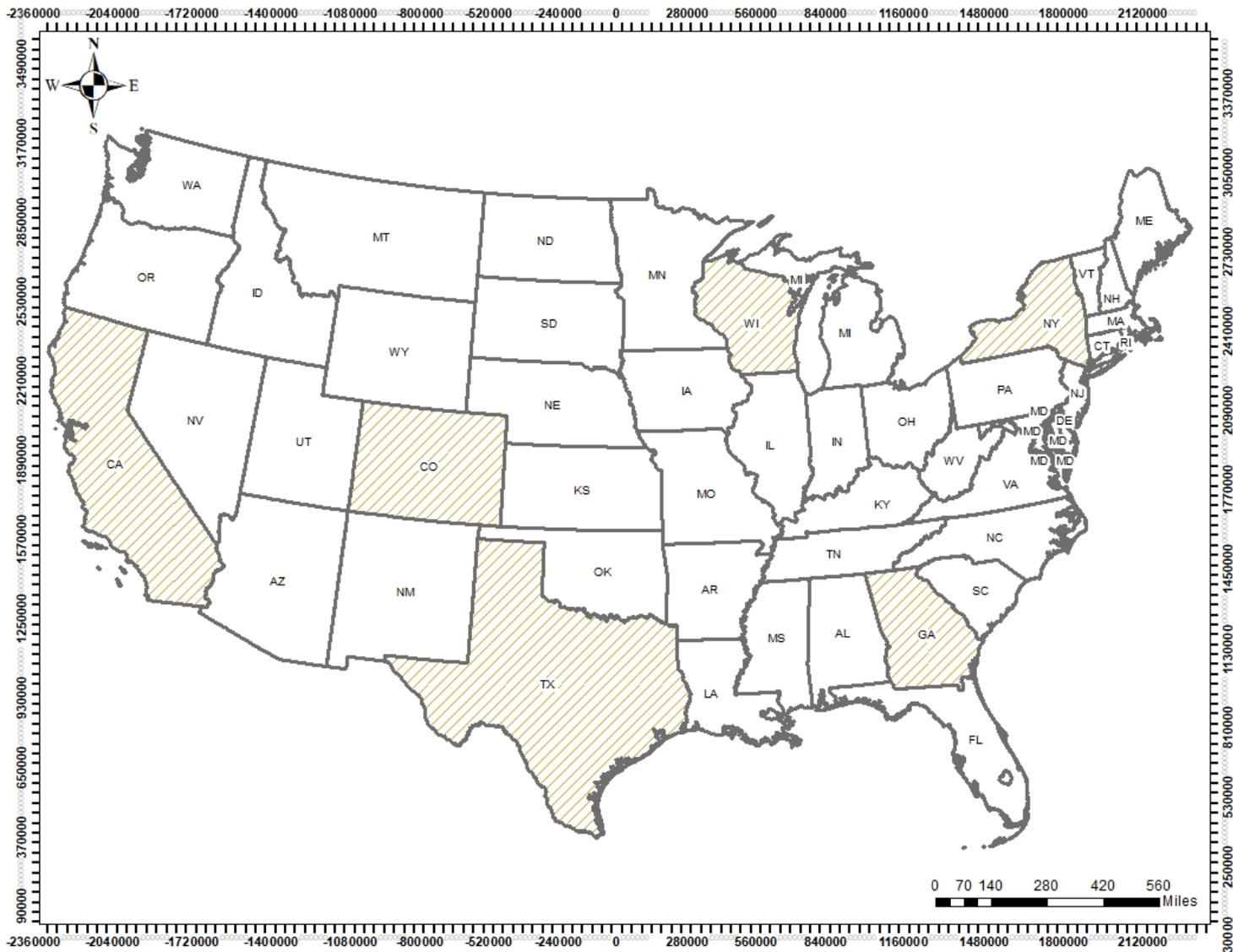
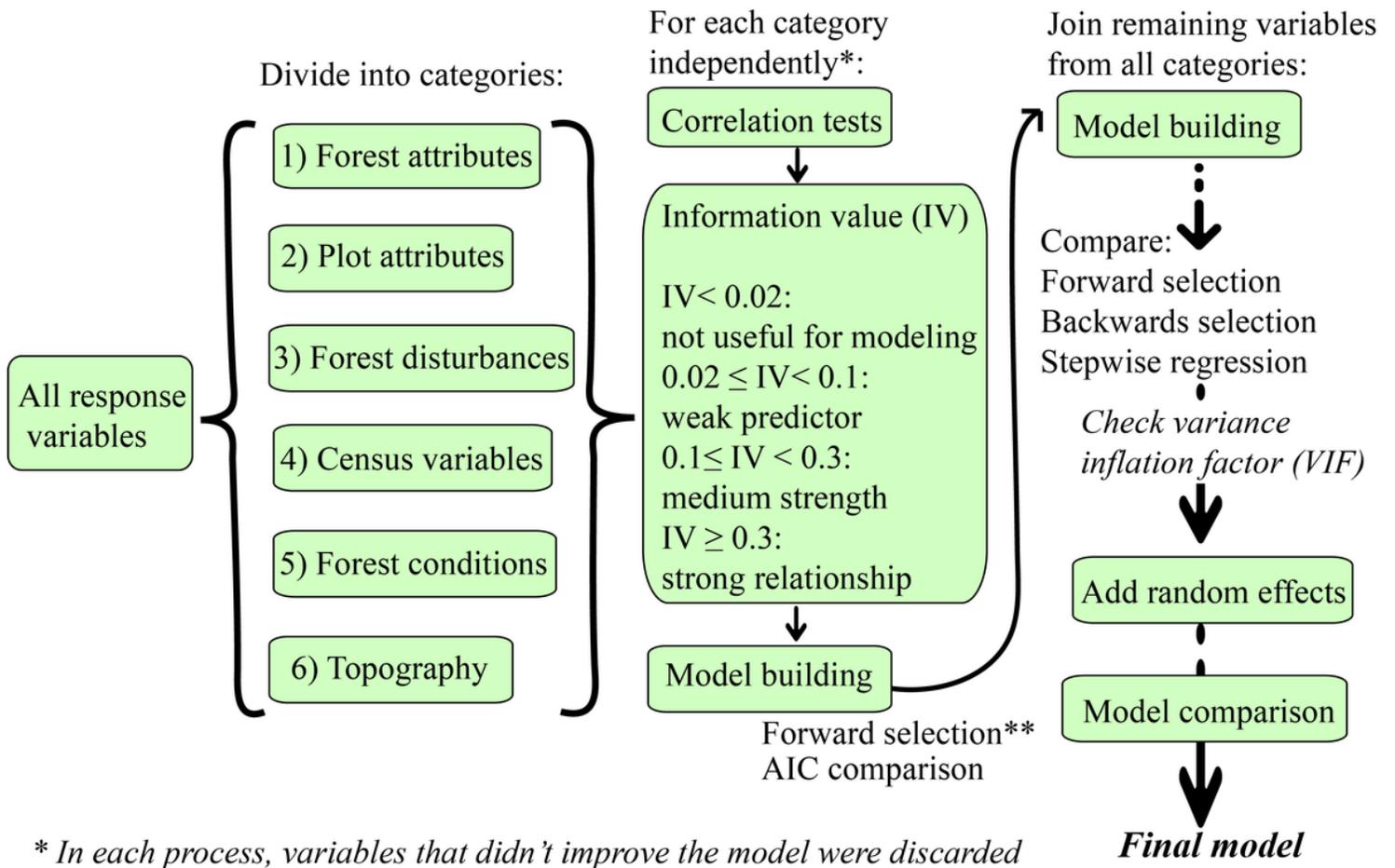


Figure 1

Study area across the United States displaying state boundaries. Highlighted areas show the states selected for this study. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



* In each process, variables that didn't improve the model were discarded

**Input order for the variables was informed by the IV

Figure 2

Workflow for the variable selection process to determine the probability of forests becoming non-forested (logistic model) and aboveground C stock changes (linear mixed-effects model).

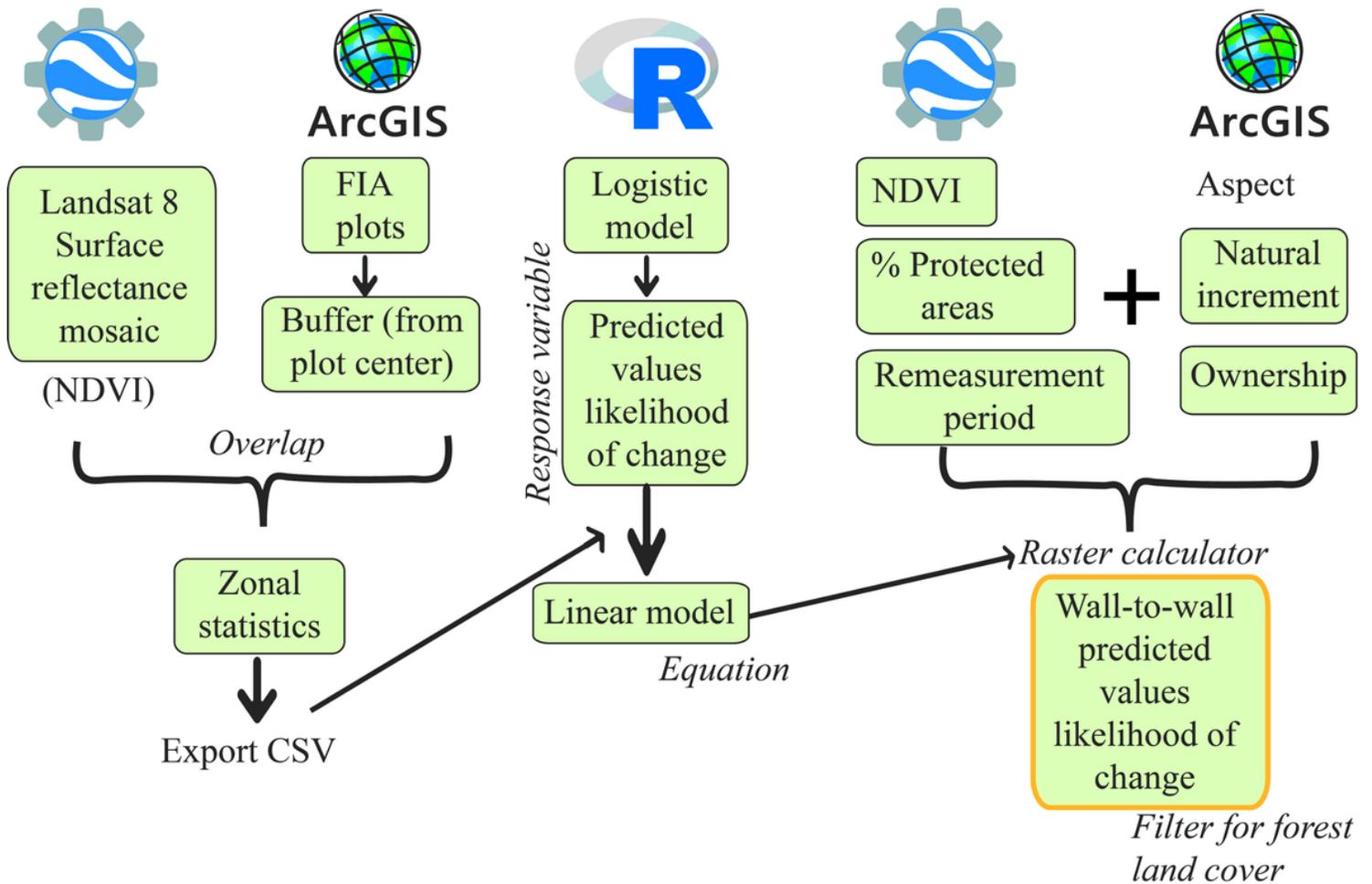


Figure 3

Step-by-step workflow for calculating the wall-to-wall likelihood of forest land use change for the study area showing processes in both ArcGIS and R environments and data layers used in each.

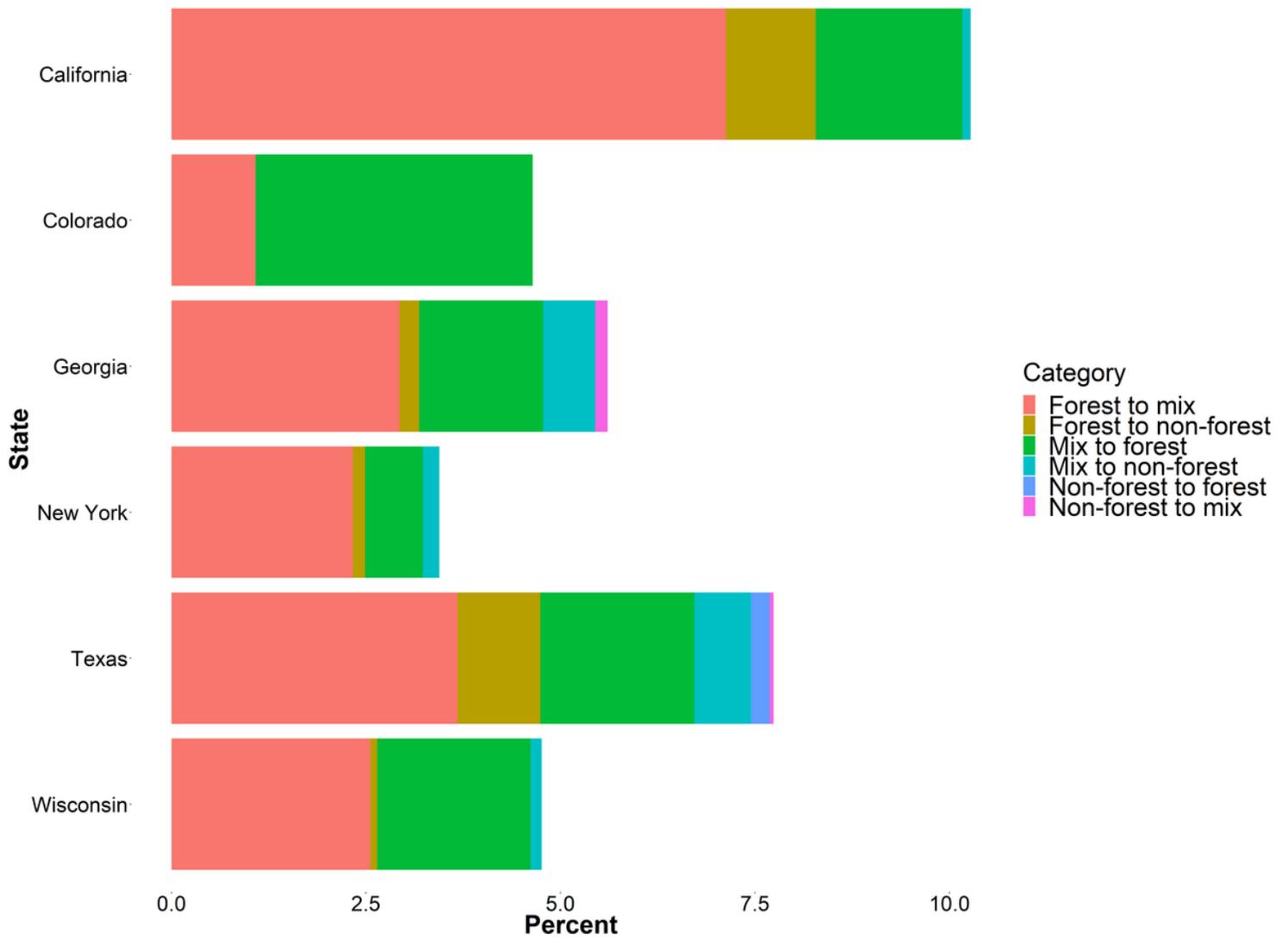


Figure 4

Percent of FIA plots across six US states that have transitioned to and from forested conditions in the study area between 2000 - 2017.

Fig. 5a

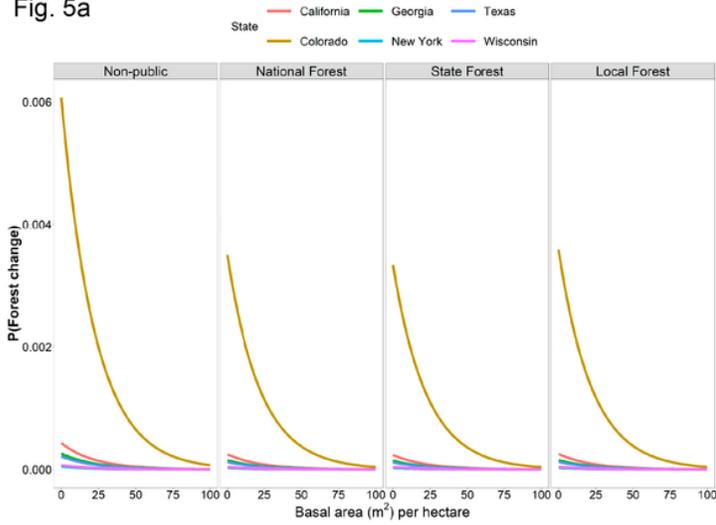


Fig. 5b

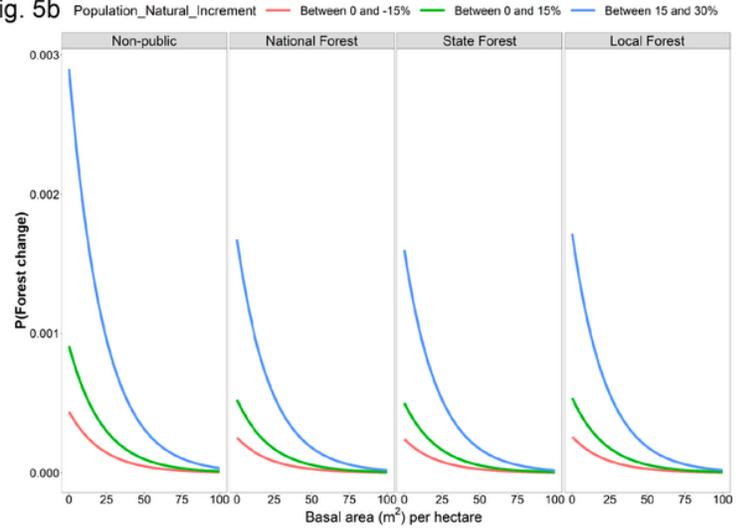


Fig. 5c

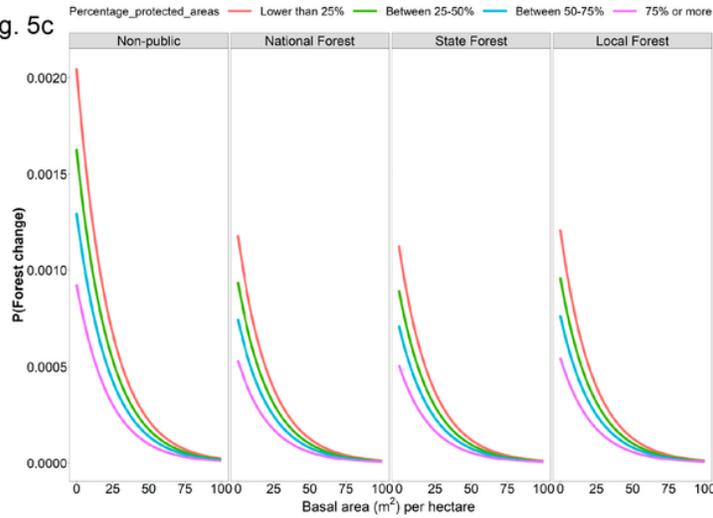


Figure 5

Predictions from the logistic regression model. Response variable is the probability of forest land use change to mixed or non-forest conditions. The four horizontal panels show different land ownership classes. Fig. 5a. The probability of land use conversion for the six states analyzed. Fig. 5b. The effects of population natural increment. Fig. 5c. The effects of the percentage of protected areas in a county.

Fig 6a

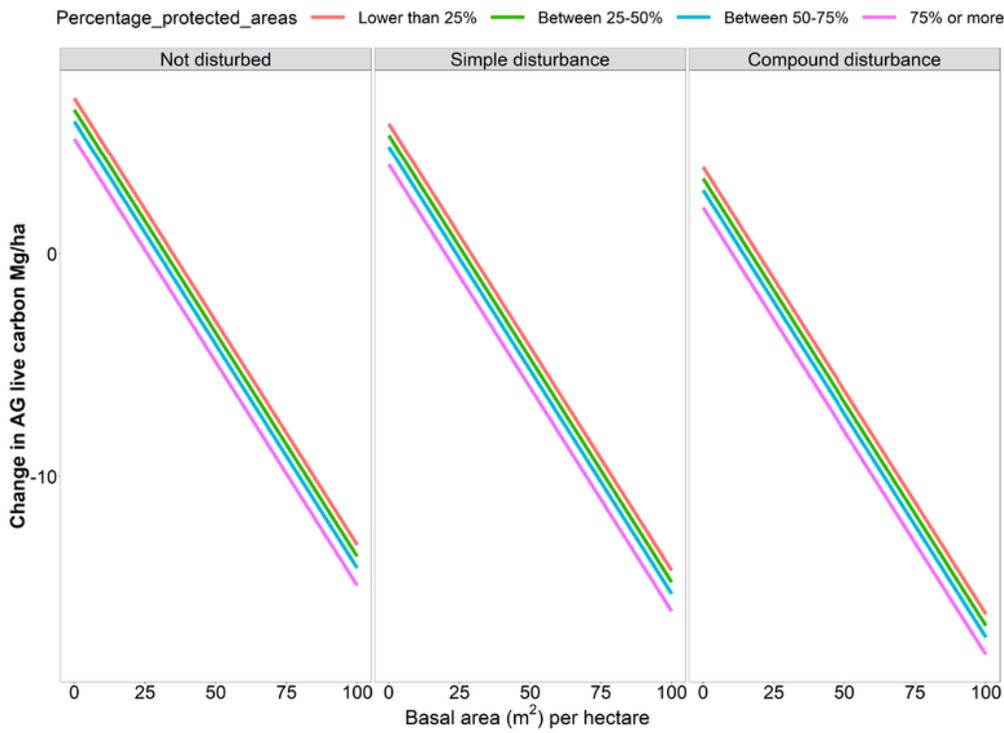


Fig 6b

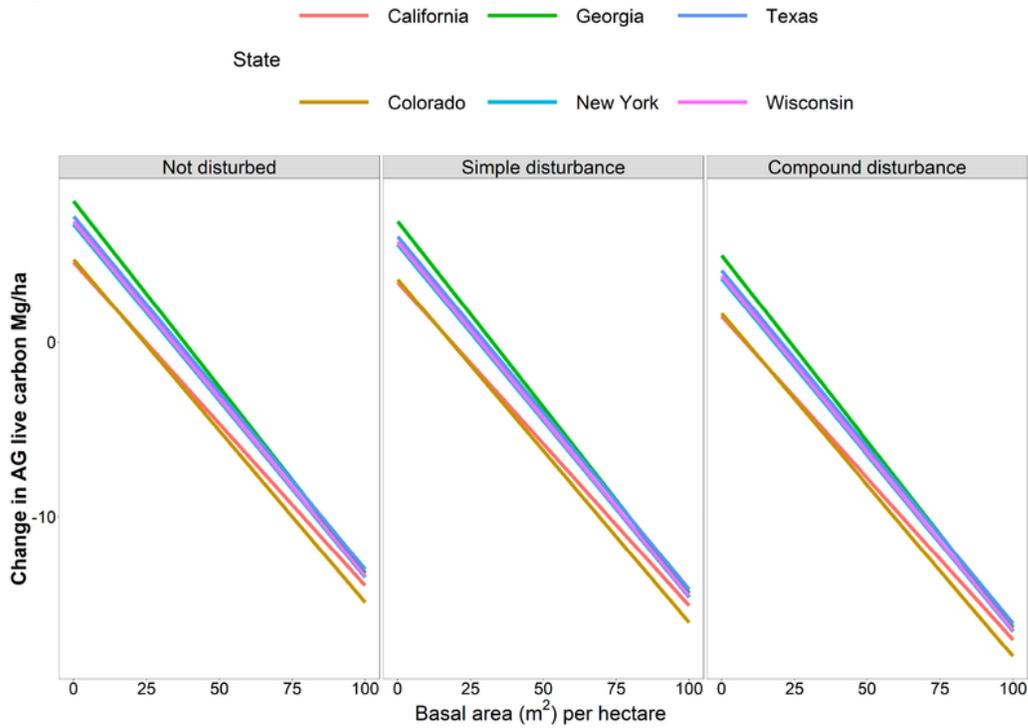


Figure 6

Predictions showing the change in aboveground (AG) live C stocks. The three horizontal panels show different disturbances classes. Fig. 6a. Shows the change in C for the six states analyzed. Fig 6b. Shows the effects of the percentage of protected areas in a county.

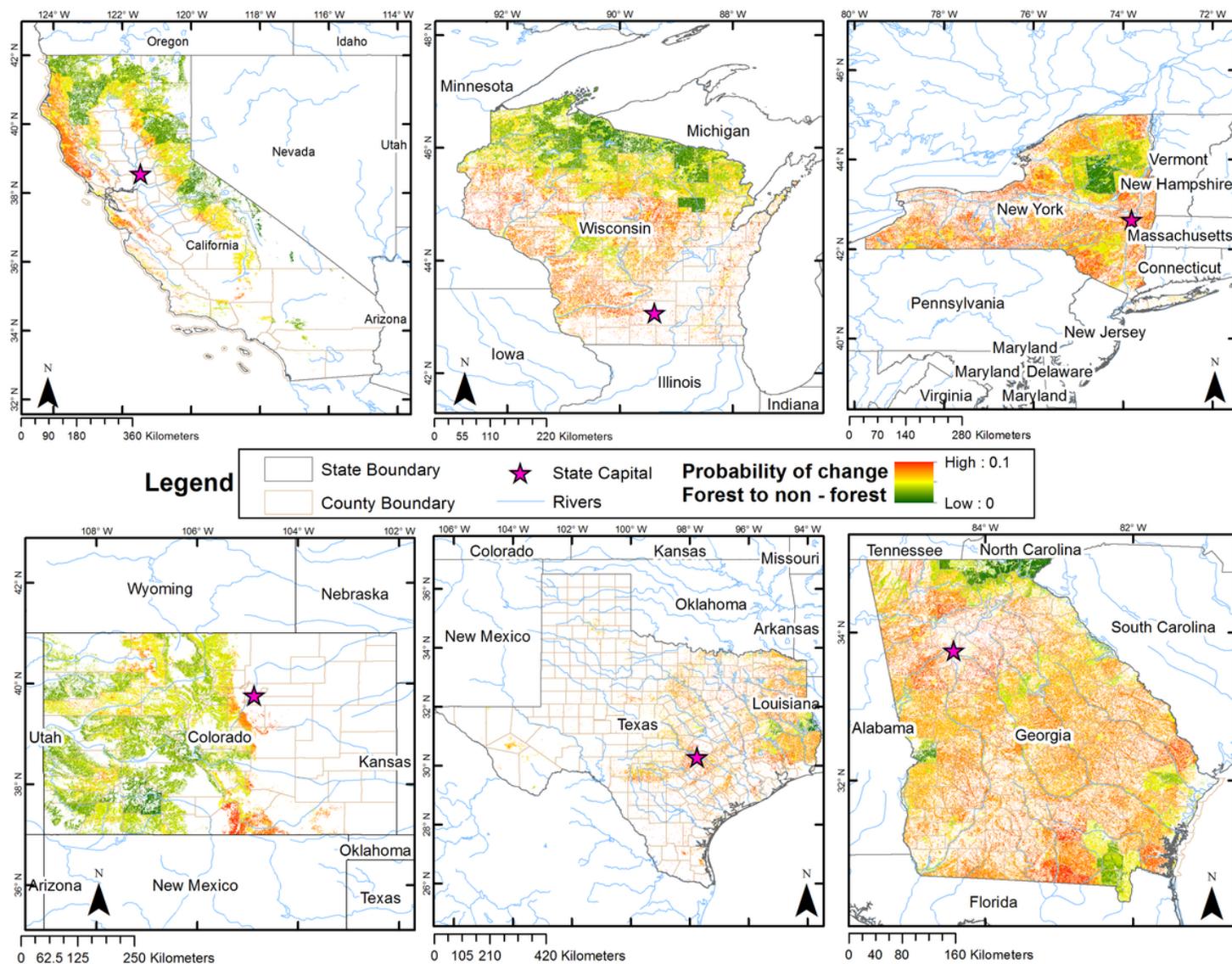


Figure 7

Wall-to-wall maps of the probability of forest areas to convert to non-forest. Greener areas show less probability of conversion, while yellow areas represent a moderately low probability, orange areas represent a moderately high probability, and dark orange-red colors show a higher probability of conversion. Probability values range from 0 to 0.1 Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)

- [Additionalfile2.docx](#)
- [Additionalfile3.docx](#)