Science of the Total Environment 857 (2023) 159252

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Empirical nitrogen and sulfur critical loads of U.S. tree species and their uncertainties with machine learning

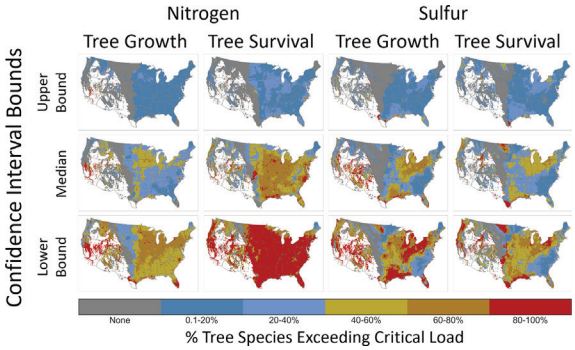
Nathan R. Pavlovic [a,*], Shih Ying Chang [a], Jiaoyan Huang [a], Kenneth Craig [a], Christopher Clark [b], Kevin Horn [c], Charles T. Driscoll [d]

[a] Sonoma Technology, 1450 N. McDowell Blvd., Suite 200, Petaluma, CA 94954, United States
[b] U.S. Environmental Protection Agency, Office of Research and Development, 1200 Pennsylvania Avenue, NW, Mail Code: 8101R, Washington, DC 20460, United States
[c] Data and Environmental Scientist, Freedom Consulting Group, 7061 Columbia Gateway Drive, Columbia, MD 21046, United States
[d] Department of Civil and Environmental Engineering, Syracuse University, 151 Link Hall, Syracuse, NY 13244, United States

## HIGHLIGHTS

- We use machine learning modeling to quantify critical loads and their uncertainty.
- Machine learning models performed better than other modeling techniques.
- The uncertainty is as large as deposition variability for many tree species.
- Uncertainty in our results is relevant to decision making and air quality standards.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

Editor: Elena Paoletti

*Keywords:*
Nitrogen deposition
Sulfur deposition
Tree growth
Tree survival

## ABSTRACT

Critical loads (CLs) of atmospheric deposition for nitrogen (N) and sulfur (S) are used to support decision making related to air regulation and land management. Frequently, CLs are calculated using empirical methods, and the certainty of the results depends on accurate representation of underlying ecological processes. Machine learning (ML) models perform well in empirical modeling of processes with non-linear characteristics and significant variable interactions. We used bootstrap ensemble ML methods to develop CL estimates and assess uncertainties of CLs for the growth and survival of 108 tree species in the conterminous United States. We trained ML models to predict tree growth and survival and characterize the relationship between deposition and tree species response. Using four statistical methods, we quantified the uncertainty of CLs in 95 % confidence intervals (CI). At the lower bound of the CL uncertainty estimate, 80 % or more of tree species have been impacted by nitrogen deposition exceeding a CL for tree survival over >50 % of the species range, while at the upper bound the percentage is much lower (<20 % of tree species impacted across >60 % of the species range). Our analysis shows that bootstrap ensemble ML can be effectively used to quantify critical loads and their uncertainties. The range of the uncertainty we calculated is sufficiently large to warrant consideration in management and regulatory decision making with respect to atmospheric deposition.

## 1. Introduction

Emissions of sulfur oxides ($SO_x$), nitrogen oxides ($NO_x$), and ammonia ($NH_3$) result in atmospheric deposition of nitrogen and sulfur compounds

* Corresponding author.
*E-mail address:* npavlovic@sonomatech.com (N.R. Pavlovic).

http://dx.doi.org/10.1016/j.scitotenv.2022.159252
Received 3 August 2022; Received in revised form 29 September 2022; Accepted 1 October 2022
Available online 8 October 2022

onto the land surface, which can lead to acidification and eutrophication (nitrogen enrichment) of terrestrial and aquatic ecosystems (Fenn et al., 2011). The cumulative impacts of atmospheric deposition on ecosystems impact timber production, carbon sequestration, soil fertility, biodiversity, habitat preservation, commercial and recreational fishing, and tourism. The U.S. Environmental Protection Agency (EPA) has found that current secondary National Ambient Air Quality Standards (NAAQS) for $NO_2$ and $SO_2$ are "not adequate to provide appropriate protection against deposition-related effects associated with oxides of nitrogen and sulfur" (U.S. Environmental Protection Agency, 2011). The Clean Air Act grants EPA the authority to establish secondary standards to protect public welfare, but currently the secondary standards are set to the same level as the primary standards. In addition to the EPA and the Clean Air Act, other agencies, including the U.S. Forest Service and National Park Service, have regulatory authority to address the impact of poor air quality on land management objectives under the Federal Lands Policy and Management Act, the National Environmental Policy Act, the Wilderness Act, and the National Forest Management Act. Agency decision-making commonly uses data and modeling results indicating the potential impact of atmospheric deposition on ecological endpoints.

In the context of atmospheric deposition, critical loads (CLs) are estimates of a quantitative threshold for levels of deposition above which significant harmful ecological effects occur (Phelan et al., 2018; Nilsson and Grennfelt, 1988; Burns and Sullivan, 2015). The CL concept was developed through the Convention on Long-Range Transboundary Air Pollution (CLRTAP, 2004). EPA uses estimates of CLs to guide review of the secondary NAAQS for ecological effects of $NO_x$ and $SO_x$ under the Clean Air Act (U.S. Environmental Protection Agency, 2017). Multiple methods have been reported for estimation of CLs, including empirical methods, steady state mass balance (e.g., McNulty et al., 2007), and dynamic modeling (Wu and Driscoll, 2010; Blett et al., 2014).

CLs for different taxonomic groups such as trees (Horn et al., 2018), herbaceous communities (Simkin et al., 2016; Clark et al., 2019), and lichen (Geiser et al., 2019, 2010) are commonly determined using empirical relationships developed through observational data (Pardo et al., 2011). The empirical relationship is typically described in the form of a dose-response function. CLs have been considered the point along the dose-response relationship at which a relevant negative biological response occurs (Pardo et al., 2019), and we use the term in this sense throughout. Researchers have defined this threshold using several approaches. For epiphytic macrolichens, Geiser et al. (2019) define the CL as the condition at which a 20 % decline in response is observed. Simkin et al. (2016) developed quantile regression models of the richness and estimated N CLs of herbaceous plant species using the partial derivative of the regression model. Clark et al. (2019) conducted binomial generalized linear models for each individual herbaceous species by examining their response to environmental data (S deposition, N deposition, soil pH, precipitation, and temperature). Nitrogen CLs were calculated by taking the partial derivative of the best statistical model with respect to N and S deposition and solving for deposition. Horn et al. (Horn et al., 2018) developed statistical relationships between atmospheric S and N deposition, and the growth and survival of 94 tree species in the conterminous United States by fitting parametric regression equations using maximum likelihood. CLs have subsequently been derived from these curves based on the point of downward inflection or maximum value of the curve. The selection of one definition or threshold approach or another can substantially influence the CL value.

There are a number of uncertainties associated with estimates of CLs (Blett et al., 2014). Biological receptors can have a range of CL values that differ because of response metrics, location, time, and environmental covariates. The vulnerability of ecosystems is a complex function of bedrock geology, climate, vegetation type, community composition, elevation, soil characteristics, ancillary disturbance, and other factors. Models that do not include interaction terms are incapable of accounting for factors that influence the sensitivity of biological receptors to atmospheric deposition. There is also substantial uncertainty in the calculation of atmospheric deposition (Walker et al., 2019). As a result, the dose-response relationships used to develop CLs may not fully represent the natural process they are intended to depict.

Because of variability in biological receptors studied, differences in methods, and the uncertainty in modeled relationships, many CLs are reported as ranges, which can vary by a factor of two or more for a given region, ecosystem, and biological receptor (Pardo et al., 2011). Improving the characterization, quantification, and communication of uncertainty associated with CL estimates—and reducing those uncertainties—will be essential to further develop and refine national-scale CL data to guide regulatory action (Blett et al., 2014; Clark et al., 2019). Modeling techniques that address the limitations of previous modeling approaches to develop dose-response curves, such as through improved model accuracy and representation of variable interactions, can help better quantify and reduce uncertainty.

ML models are well-suited to address the limitations of other methods used in CL analysis. ML models have among the highest levels of accuracy in modeling empirical relationships (Lary et al., 2016). Furthermore, ML models generally do not rely on the parametric assumptions of the form of relationships between variables, which allows for greater flexibility in the form of those relationships (i.e., linear, step-function, and polynomial relationships may all be represented with the same model form). In addition, decision tree type models inherently represent and can be used to describe variable interactions. The XGBoost ML algorithm (Chen and Guestrin, 2016) in particular has demonstrated excellent predictive performance in competitions of ML accuracy (Bekkerman, 2015) and environmental sciences (Ma et al., 2019). As a result, the algorithm has gained widespread use in predictive tasks. Recent advances in model interpretation techniques offer the opportunity to identify key variable relationships within ML models (Molnar, 2019) through causal inference (Zhao and Hastie, 2021). Limitations of ML models include the requirement of substantial training data, significant computational requirements, and the potential to replicate bias within the training set through overfitting. Nevertheless, ML techniques have significant advantages and have never been used in the assessment of CLs. Therefore, they may provide a powerful tool to examine complex relationships that have been difficult to capture with other methods.

In this paper, we introduce a novel approach to estimate CLs derived from ML models for a collection of over 100 tree species across the conterminous United States. We hypothesized that machine learning modeling would provide improved performance for prediction of tree outcomes relative to previous modeling efforts and that the model results could be used to quantify critical load uncertainty. We build on Horn et al. (Horn et al., 2018), who used simulated annealing to fit parameters for an a priori determined functional form (i.e., multiplicative product of Gaussian and power functions) to relate the growth and survival of tree species to six different factors (i.e., N deposition, S deposition, temperature, precipitation, tree size, and competition among neighbors [calculated from the plot basal area and the basal area of trees larger than that of the tree observed tree]). Using these fitted models, researchers have since been able to estimate CLs for N and S for individual tree species in exploratory analyses. We used XGBoost with the remeasured tree inventory data to predict tree growth and survival, and we used the relationships represented in the trained models to develop empirical relationships between atmospheric deposition and growth and survival of individual tree species. We used four different statistical tests to quantify (1) the CL of the growth and survival of a tree species, and (2) the associated uncertainty of the CL estimator ($\widehat{CL}$) with bootstrapping (Efron, 1979). We compared these results to CLs derived from the relationships developed by Horn et al. (2018) and assessed the uncertainty of these CLs based on the bootstrapped 95 % confidence interval (CI). We further assessed the extent to which these CLs have been exceeded historically across the spatial ranges of the tree species.

## 2. Materials and methods

### 2.1. Forest inventory data

To facilitate a comparison with other CL methods, we obtained tree growth and survival data sets for individual tree species used by Horn

et al. (2018). Briefly, the data set was compiled from the Forest Inventory and Analysis (FIA) program database of the U.S. Forest Service in January 2017 covering data for 2000–2016. Growth and survival for individual tree species were calculated using the first and last inventory observations available for each tree in the database. The data set included indices of tree height, basal area, and above-ground biomass from FIA. Environmental conditions at each site were added to the data set using (1) monthly mean temperature and precipitation data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) Climate Group at Oregon State University (23, 24) and (2) atmospheric deposition of S and N data from U.S. National Atmospheric Deposition Program's Total Deposition Science Committee (TDep) (Program, 2021; Schwede and Lear, 2014). 1,260,622 trees were included in the growth data set and 1,544,523 trees in the survival data set. Species with 1500 or more observed trees were used to develop ML models for tree growth and survival. This cutoff was selected to allow calculation of CLs for additional tree species while balancing higher uncertainty for tree species with fewer observations. Atmospheric S and N deposition, tree height, basal area, above-ground biomass, and the number of years between the first and last observation were used as features (predictors) for model development. We then developed an updated forest inventory data set based on the January 2021 FIA database, using the same methods as in the Horn et al. (Horn et al., 2018) data set.

### 2.2. ML model

We developed growth and survival models for 108 tree species using the XGBoost ML algorithm (Chen and Guestrin, 2016). XGBoost is a tree-based ML ensemble algorithm in which weak prediction models, in the form of classification and regression (CART) trees, are iteratively trained to correct the residuals of previous models. To avoid confusion between statistical trees (i.e., CARTs) and botanical trees, we hereafter refer to the former as "CARTs" and the latter as "trees." We developed models to predict (1) annual growth rate and (2) decadal survival probability for different tree species, which were used to develop empirical CLs for both N and S. We trained these models using full random search hyperparameter tuning, using a 75 % training set, and evaluated them using the remaining 25 % of the data. We define the input data set, which is composed of $n$ records and $m$ features, as $D = \{(x_i, y_i)\}(|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$, where $x_i$ is the $i^{th}$ set of predictive features input to the model and $y_i$ is the $i^{th}$ value to be predicted. Our model structure is

$$\widehat{y}_i = \widehat{f}(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{1}$$

where $\widehat{y}_i$ is the predicted tree growth or survival, and K is the number of additive functions making up the XGBoost ensemble. $f_k(x_i)$ is the $k^{th}$ independent function, where $f_k$ is an individual CART selected from the space $F = f\{(x) = w_{q(x)}\}(q : R^m \to T, w \in \mathbb{R}^T)$ for model leaf weights $w$ of independent CART structure $q$ composed of T leaves from $m$ input data features. CART structures were iteratively trained using gradient boosting, as described by Chen and Guestrin (2016). For the tree survival model, we use the negative log loss objective function.

$$\textit{negative log loss} = -\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j} \log\left(p_{i,j}\right) \tag{2}$$

where N is the number of samples, M is the number of classes (live or dead), $y_{i,j}$ is a binary indicator of whether observation i belongs to class j, and $p_{i,j}$ is the predicted class probability. For the tree growth model, we train the model using the mean square error objective function.

$$\textit{mean square error} = \sum_{i=1}^{N} (y_i - \mathrm{f}(x_i))^2. \tag{3}$$

We assessed model performance using $R^2$ (Kvålseth, 1983), root mean square error (RMSE), and mean absolute error (MAE) for growth, and, for survival, the receiver operating characteristic's (ROC) area under the curve (AUC) (Fawcett, 2006). The AUC represents the probability that, in a two-class classification model, a member of the positive class will have a higher predicted probability of membership in the positive class than a member of the negative class.

The ML models were trained using the Scikit-learn 0.23.2 and XGBoost 1.2.0 libraries in Python 3.6.

### 2.3. Model interpretation

To investigate the relationships between atmospheric deposition and tree growth and survival, we calculated the partial dependence (PD) of the model predictions on S and N deposition. The PD function provides the marginal effect of a feature on the predicted outcome of a ML model (Friedman, 2001), while all other variables are held constant. The PD function is defined and approximated as

$$\widehat{f}_{xS}(x_s) = E_{xC}\left[\widehat{f}(x_S, x_C)\right] \approx \frac{1}{N}\sum_{i=1}^{N}\widehat{f}\left(x_S, x_C^i\right), \tag{4}$$

where $x_S$ is the feature vector for the feature S for which the partial dependence is calculated, $x_C$ is the set of feature vectors for all other predictive features C, and $x_C^i$ are the actual values of features C for the record i.

To aid in interpretation of the PD relationship result for the tree growth model and ensure comparability with the Horn et al. (Horn et al., 2018) method, we further hold the value of number of years between first and last observation constant at 10 years during calculation of the approximate PD ($\widehat{PD}$).

### 2.4. Critical load estimation

We used the approximated PD function for N and S deposition to estimate the CL for each tree species for both tree growth and tree survival. To estimate the CL for each model, we fit a polynomial curve $\widehat{PD}_{poly} = \beta_0 + \beta_1 x_S + \beta_2 x_S^2 + \beta_3 x_S^3 + \varepsilon$, up to the third degree, to the PD values calculated at even intervals within the range of observed deposition values for each tree species. The polynomial fit provides a smoothed approximation of the PD function. We defined the CL as the minimum value of $x_s$ along the N or S deposition gradient where there is a 1 % decrease from the maximum value. Specifically, for $\widehat{PD}_{poly}$ with a downward sloping component as $x_S$ increases, we defined the value of $x_S$ that maximizes $\widehat{PD}_{poly}$ as $x_{\widehat{SPD}_{max}}$. We calculated the CL as the minimum value of $x_S$ greater than $x_{\widehat{SPD}_{max}}$ for which

$$\frac{\widehat{PD}_{poly}\left(x_{\widehat{SPD}_{max}}\right) - \widehat{PD}_{poly}(x_S)}{\widehat{PD}_{poly}\left(x_{\widehat{SPD}_{max}}\right)} \geq 0.01. \tag{5}$$

We define this minimum value of $x_S$ as the $\widehat{CL}_{poly}$. A graphical representation of this value is illustrated in Fig. S4a. Note that in this effort we define the 1 % reduction level as the CL. Relatively small demographic rate changes can alter forest structure and function due to compounding effects over time (Kobe, 1996) We selected the threshold used here as impactful to forest management objectives due to a significant loss of function, even at a relatively small absolute change (van Mantgem et al., 2009). This is similar in spirit to the lichen CLs (Geiser et al., 2019) where a 20 % change in the community is defined as the threshold and is different from Horn et al. (Horn et al., 2018). Horn et al. (2018) did not formally define the estimates as CLs, but it is implied that the statistical terms therein could be interpreted as CLs.

### 2.5. Comparison to other critical load estimates

To assess differences between the ML approach and previously reported methods, we compared the model performance to CLs calculated using the

regression equations described by Horn et al. (2018). CLs were defined as the deposition value at which 1 % decline from the maximum growth or survival occurred in the Horn et al. (2018) regression equations. For the growth regression model, we also compared model performance using model $R^2$. This comparison was made using $\widehat{CL}_{poly}$ generated from the entire, original training data set rather than the bootstrapped data sets described in Section 2.6. This assessment was performed to provide directly comparable results to the Horn CLs based on the same input data sets.

### 2.6. Uncertainty assessment

To address uncertainty in ML-based CL values created using $\widehat{PD}_{poly}$, we developed a bootstrap ensemble of XGBoost models. Six hundred XGBoost models were trained on a resampling of the original data set of size equal to the original data, randomly sampled with replacement. We fit a function $\widehat{PD}_{poly,b}$ to each bootstrap model $b$ and defined the 95 % CI as the range representing the 2.5th and 97.5th percentiles of $\widehat{CL}_{poly,b}$ across all bootstraps for which the $\widehat{CL}_{poly,b}$ could be calculated (Fig. S4b).

Using the bootstrap results, we further developed CL estimates that do not depend on polynomial approximation of $\widehat{PD}$. To develop uncertainty-based CLs, we first test whether tree growth and survival are significantly related to deposition. To test this condition, we calculate the 95 % CI for $\widehat{PD}$ at each of 20 bins that are evenly spaced between the minimum and maximum observed atmospheric S and N deposition for the geographic range of occurrence of each tree species. The 95 % CI is 2.5th and 97.5th percentiles of $\widehat{PD}$ across all bootstraps. For each tree species, we first test whether a significant relationship exists between N or S deposition and tree growth or survival by calculating

$$\max\left(\left|\widehat{PD}\left(x_{\widehat{SPD}_{max}}\right)-\widehat{PD}(x_S)\right|\right) \tag{6}$$

for all values for $x_S > x_{\widehat{SPD}_{max}}$ for each bootstrap. If the bootstrapped 95 % CI for this value does not include zero, we conclude a significant relationship exists between deposition and tree growth and survival. If a significant relationship exists, we next calculate the minimum value of $x_S$, for which, for all $x_S > x_{\widehat{SPD}_{max}}$, the 95 % CI of $\widehat{PD}(x_S)$ does not overlap the 95 % CI of $\widehat{PD}\left(x_{\widehat{SPD}_{max}}\right)$. The minimum value that satisfied these conditions is the lowest level of deposition at which a decline in tree function is detected and is the statistically significant CL, $\widehat{CL}_{stat}$ (Fig. S4c). $\widehat{CL}_{stat}$ represents the minimum deposition level at which the predicted growth or survival from 95 % of bootstraps is below the maximum growth or survival level across 95 % of bootstraps. At this level, there is a high statistical certainty that a negative effect due to atmospheric deposition exists, but the magnitude of the effect is not considered.

It is possible that a $\widehat{CL}_{stat}$ may correspond to a very small decline of tree function. To calculate the level at which a meaningful decline in tree function occurs, we estimate the deposition level at which a 1 % decline in tree growth or survival is identified by $\widehat{PD}$ for each bootstrap $\widehat{CL}_{np}$. To estimate $\widehat{CL}_{np}$, we calculate the minimum value of $x_S$, where $x_S > x_{\widehat{SPD}_{max}}$, which satisfies

$$\frac{\widehat{PD}\left(x_{\widehat{SPD}_{max}}\right)-\widehat{PD}(x_S)}{\widehat{PD}\left(x_{\widehat{SPD}_{max}}\right)} \geq 0.01. \tag{7}$$

$\widehat{CL}_{np,b}$ was then approximated for each bootstrap as

$$0.99\left(\widehat{PD}\left(x_{\widehat{SPD}_{max}}\right)\right) * \frac{\widehat{PD}(x_{S-1})-\widehat{PD}(x_S)}{x_{S-1}-x_S} \tag{8}$$

where $x_{S-1}$ is the first value of $\widehat{PD}$ less than $x_S$. The 95 % CI for $\widehat{CL}_{np,b}$ was then estimated from the bootstrapped values (Fig. S4d).

In some cases, CLs from one or more CL calculation methods could not be calculated for a tree species although sufficient data were available for that species. For $\widehat{CL}_{stat}$, no CL can be calculated in cases where the variability tree growth and survival between bootstraps is greater than the variability within the bootstraps. For the other methods, CLs cannot be calculated if statistical assessment shows that no relationship exists between deposition and tree growth or survival or in cases where the PD showed a flat or increasing trend.

### 2.7. Exceedance analysis

Using a 5-year average (2015–2019) of TDep atmospheric N and S deposition, we assessed the extent to which recent deposition has exceeded the CLs for different tree species at different confidence levels. We generated maps indicating whether observed deposition was higher than the 1) lower bound, 2) median, or 3) upper bound of the 95 % CI for $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ bootstrap results for each tree species across the published range for that species (Little, 1971, 1976, 1977, 1978). To summarize the extent of exceedance across the U.S., we calculated the proportion of all species present in each location for which deposition exceeded the CL threshold for that species.

## 3. Results

### 3.1. Model results

Out of 90 tree species for which sufficient growth data were available, we were able, using a polynomial fit of the ML partial dependence ($\widehat{PD}_{poly}$), to estimate N CLs ($\widehat{CL}_{poly}$) for 54 species and S $\widehat{CL}_{poly}$ for 83 species using the entire data set (no bootstrap). Out of 108 species for which sufficient survival data were available, we calculated N $\widehat{CL}_{poly}$ for 78 species and S $\widehat{CL}_{poly}$ for 91 species. $\widehat{CL}_{poly}$ could not be calculated for the remaining species because a 1 % drop from the maximum deposition value was not observed in the fit polynomial curve. This occurred in cases where the curve was nearly flat or had increasing deposition. Based on the independent testing set, we found that the model performance ($R^2$) for growth of tree species with variation in N and S deposition ranged from 0.2 to 0.75, with a median value of 0.54. Our model performance AUC for tree survival models with N and S deposition ranged from 0.61 to 0.92, with a median value of 0.76.

We were able to calculate bootstrapped CLs using polynomial fit $\left(\widehat{CL}_{poly,b}\right)$ and non-parametric ($\widehat{CL}_{np,b}$) methods for 69 (N, growth), 53 (N, survival), 79 (S, growth), and 68 (S, survival) tree species. Using the statistically significant decline in tree growth and survival across bootstraps ($\widehat{CL}_{stat}$), we were able to calculate CLs for 18 (N, growth), 21 (N, survival), 45 (S, growth), and 45 (S, survival) tree species. Bootstrap-based CLs were calculated only for tree species where a statistically significant relationship between deposition and tree growth or survival was identified (Eq. (6)). The bootstrapped CL estimates for all tree species for N and S are shown in Fig. 1 (for N) and Fig. 2 (for S). The $\widehat{CL}_{poly,b}$ values are derived from the $\widehat{PD}_{poly,b}$ curves (Eq. (4)) and corresponding 95 % CIs for each tree species. The deposition response curves and CIs developed for each species' growth and survival are provided in Fig. S5.

The CLs and CIs calculated with the three bootstrap methods indicate the range of likely values for the CL. For example, the CL for atmospheric deposition of S for paper birch (*Betula papyrifera)* growth ranged from 0.78 to 3.24 kg $ha^{-1}$ $yr^{-1}$ across all tests ($\widehat{CL}_{poly,b}$, $\widehat{CL}_{stat}$, and $\widehat{CL}_{np,b}$). This CL indicates one of the narrowest absolute CIs for N or S, but it nevertheless exhibits a high range of variability in terms of the percent of the median. In contrast, the confidence interval of the CL for atmospheric deposition of S for black cherry (*Prunus serotina*) growth ranged from 4.0 to 34.25 kg $ha^{-1}$ $yr^{-1}$ for the $\widehat{CL}_{poly}$ method, indicating a very wide range of absolute uncertainty.

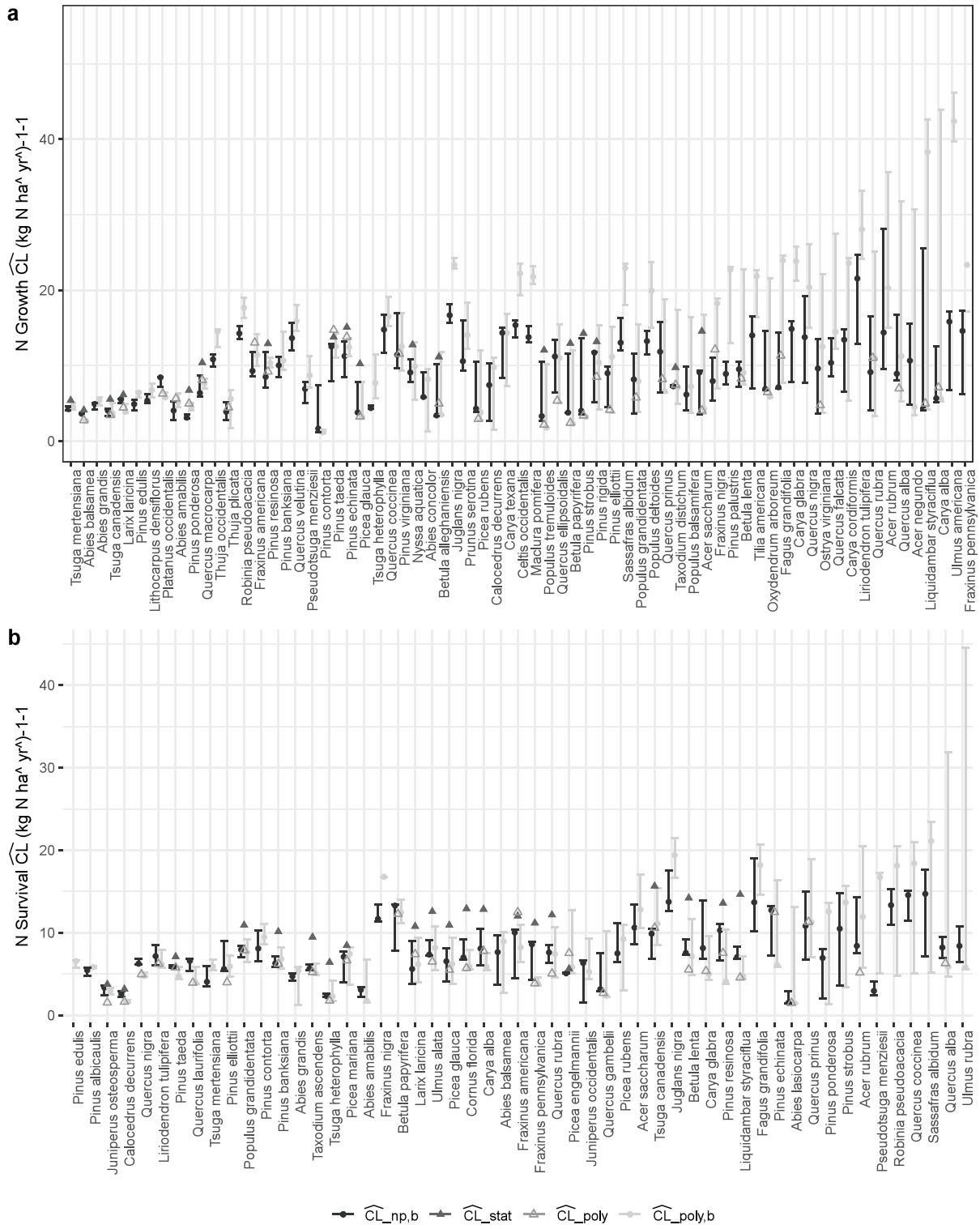



**Fig. 1.** CLs and their uncertainties for atmospheric deposition of N calculated from XGBoost bootstraps for tree growth (a) and survival (b).

We assessed the width of the CI from each technique. For N impacts on growth, the median CI width was 7.56 kg ha$^{-1}$ yr$^{-1}$ for $\widehat{CL}_{poly,b}$ and 5.05 kg ha$^{-1}$ yr$^{-1}$ for $\widehat{CL}_{np,b}$. The median values for N impacts on survival and S impacts on growth and survival were 5.22 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{poly,b}$) and 3.69 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{np,b}$), 2.16 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{poly,b}$) and 1.73 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{np,b}$), and 1.38 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{poly,b}$) and 1.66 kg ha$^{-1}$ yr$^{-1}$ ($\widehat{CL}_{np,b}$), respectively. $\widehat{CL}_{poly,b}$ CIs were wider than $\widehat{CL}_{np,b}$ CIs for for S and N impacts on growth and N impacts on survival (Fig. S1). Just over half (50.7 %) of tree species for which one or more CLs were calculated had a CL range >10 kg N ha$^{-1}$ yr$^{-1}$ when accounting for the CIs for individual methods and variation between methods for N deposition impacts on tree growth, and 29.8 % for impacts on tree survival. For S deposition, 53.2 % (growth) and 52.1 % (survival) of tree species had an overall range >5 kg S ha$^{-1}$ yr$^{-1}$.

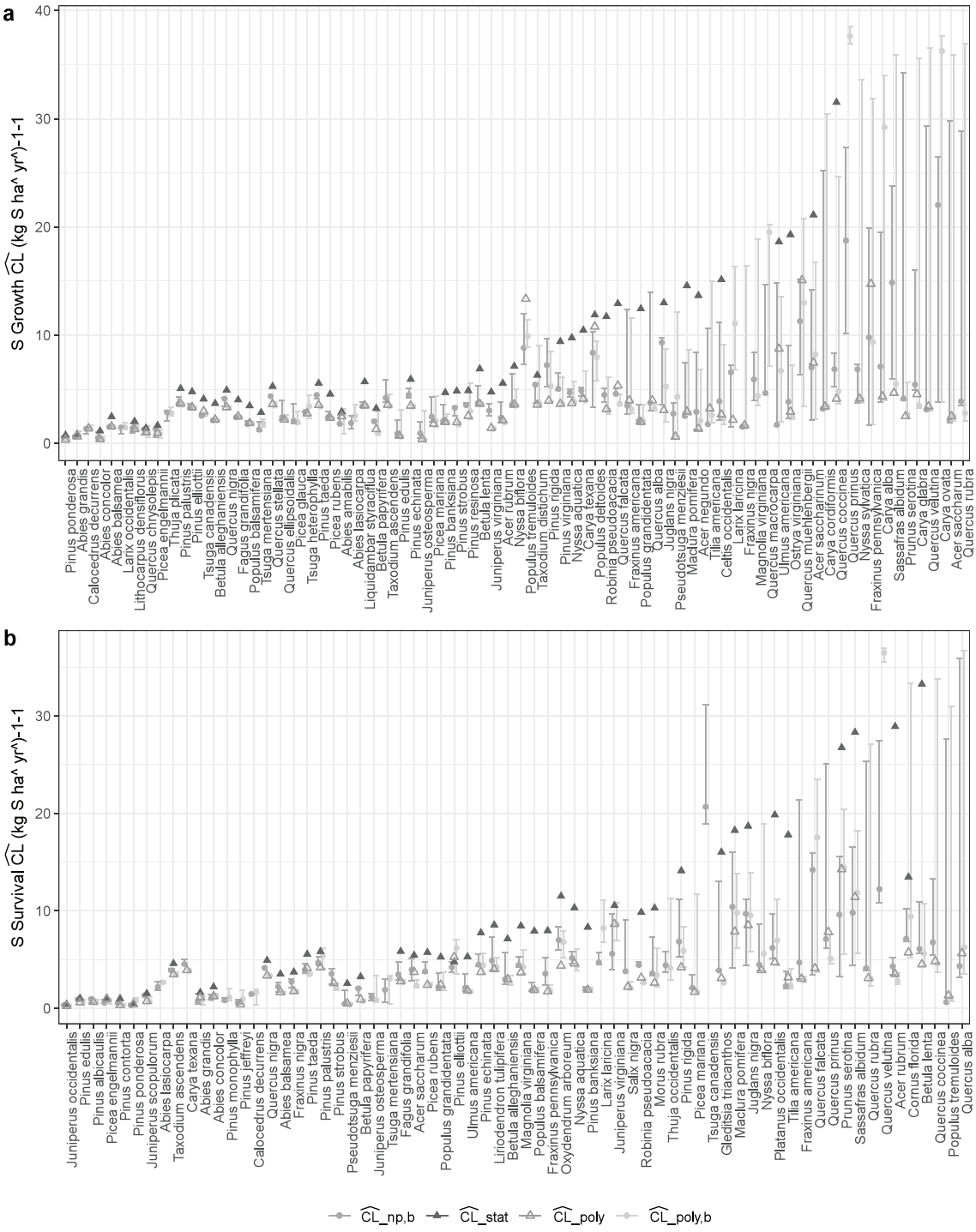



**Fig. 2.** CLs and their uncertainties for atmospheric deposition of S calculated from XGBoost bootstraps for tree growth (a) and survival (b).

The confidence intervals derived from the $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ methods show similar wide confidence intervals and medians for each species and response type, with limited exceptions (e.g., black cherry [*Prunus serotina*] growth). S $\widehat{CL}$ CIs showed greater variability in the absolute size of the CIs relative to N $\widehat{CL}$ CIs. $\widehat{CL}_{stat}$, where it could be calculated, frequently falls above the CI from the other $\widehat{CL}$ methods and provides a more conservative CL estimate. However, for a majority of tree species, no $\widehat{CL}_{stat}$ could be calculated for N deposition. This indicates that variability between bootstraps was greater than within individual bootstraps, such that a statistically significant difference between tree growth or survival over the range of observed deposition was not detected using the bootstraps.

### 3.2. Comparison with critical loads from previously reported work

Fig. 3 shows the relationship between the whole data set ML-based CLs (i.e., $\widehat{CL}_{poly}$) for growth and survival for each tree species to the 1 % reduction levels developed from the relationships defined by Horn et al. (2018).

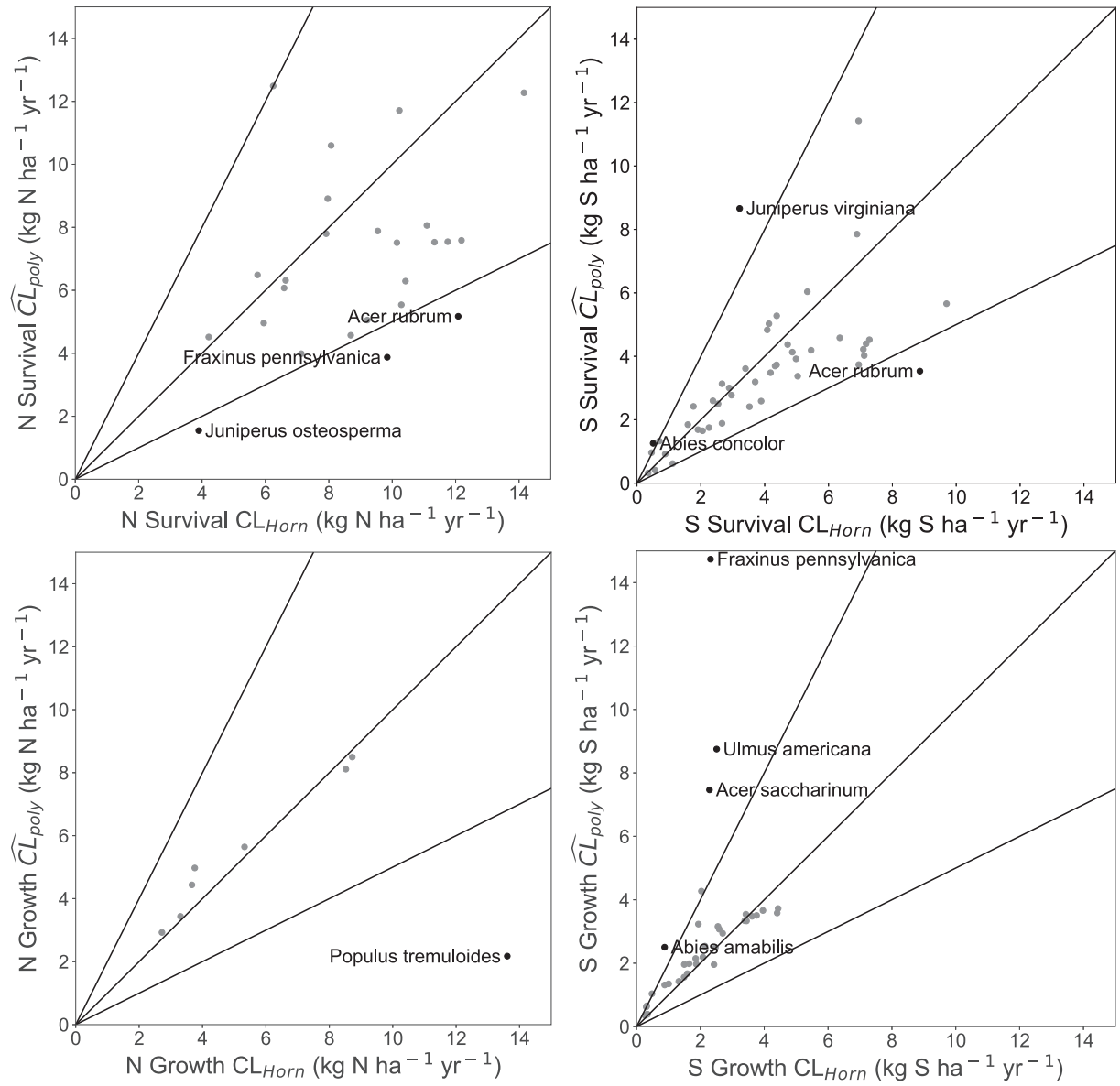



**Fig. 3.** Comparison of atmospheric N (left) and S (right) deposition 1 % reduction values reported by Horn et al. (2018) with the ML-derived $\widehat{CL}_{poly}$ values for tree survival (top) and growth (bottom). Species falling outside the 2:1 and 1:2 lines are indicated with a black point and labeled with the species name. Lower, middle, and upper solid lines indicate the 2:1, 1:1, and 1:2 lines, respectively. CLs >15 kg N ha$^{-1}$ yr$^{-1}$ are shown in Fig. S5.

Agreement was generally good between the two approaches for determining the CLs for tree survival. For tree growth, agreement was also high, but several CLs for atmospheric N deposition from ML (quaking aspen [*Populus tremuloides*], western red cedar [*Thuja plicata*], and Chestnut oak [*Quercus prinus*]) were less than half the CLs 1 % thresholds derived using Horn et al. (2018). Growth CLs for atmospheric S deposition for green ash (*Fraxinus pennsylvanica*), American elm (*Ulmus americana*), silver maple (*Acer saccharinum*), and pacific silver fir (*Abies amabilis*) from ML were more than twice the values derived using Horn et al. (2018).

For tree growth, we compared model performance ($R^2$) between the ML model and the model performance reported by Horn et al. (2018) (Fig. S2). ML model performance exceeded the performance reported by Horn et al. (2018) for all tree species, and in many cases, the $R^2$ for ML was greater by more than a factor of two. It was not possible to compare model performance for the survival results because these were not reported in Horn et al. (2018), but we would expect similar patterns.

### 3.3. Exceedance results

The proportion of tree species impacted by atmospheric deposition of N and S that exceed their CL at low, median, and high confidence are shown in Figs. 4, 5, and S3. At the median CL level within each metric, 40–60 % or more of tree species experience conditions where their CL is exceeded in some regions (i.e., the Great Lakes region for S impacts and the upper Midwest for N tree survival CLs). At the upper bound of the $\widehat{CL}$ CI, the proportion of trees exceeding the CL range for N and S is low (<20 %) across 64 to 92 % of the portion of the conterminous United States in the range of one or more tree species (Table S1). At the lower bound of the $\widehat{CL}$ CI for N deposition impacts on tree survival, the proportion of tree species exceeding the CL is high (>80 %) across 53 to 61 % of the range. For S and N growth CLs and S survival CLs, the upper bound threshold also shows an increase in the proportion of tree species exceeding the CL.

Comparison of the proportion of species exceeding the calculated CL between the $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ methods indicate differences in the results

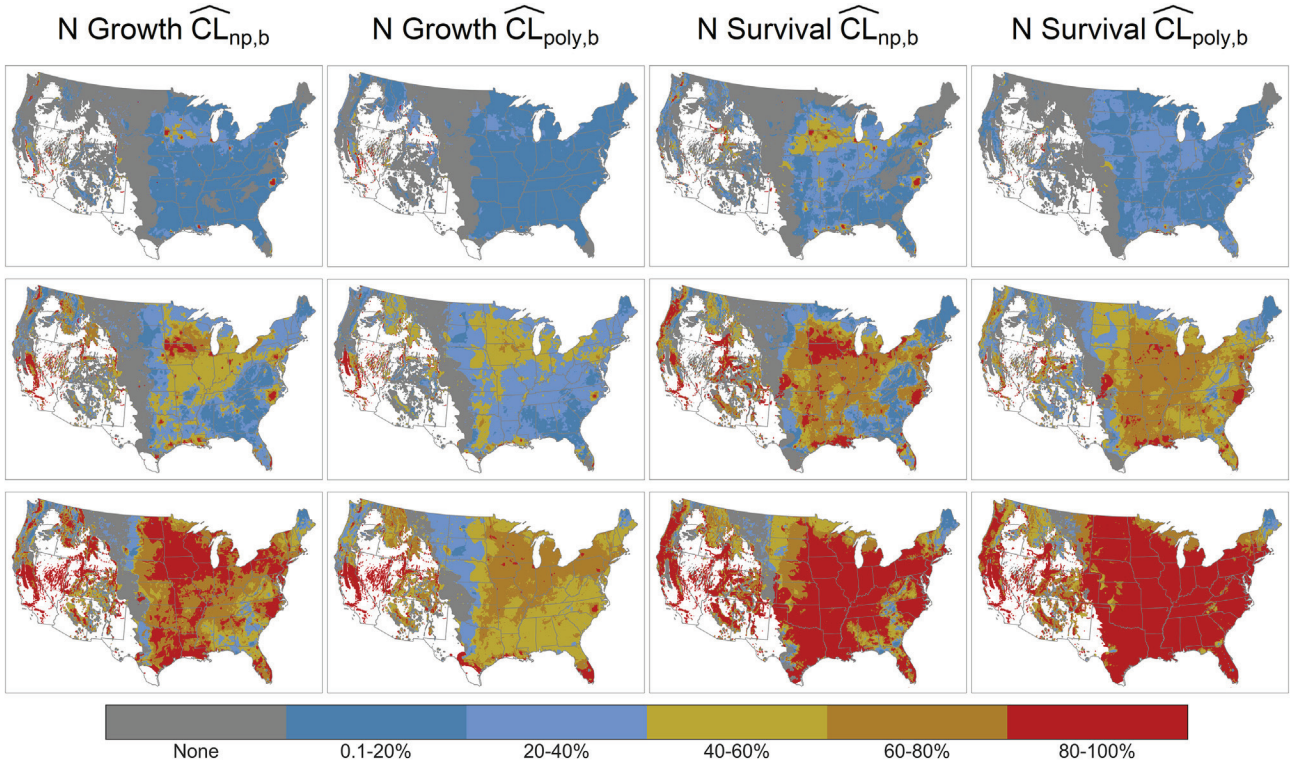



**Fig. 4.** Percentage of tree species exceeding their calculated N CL from $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ at high (upper 95th CI bound; top), median (median CL value; middle), and low (lower 95th CI bound; bottom) values within the bootstrapped CI. White indicates locations where none of the tree species for which a CL CI could be calculated are known to occur.

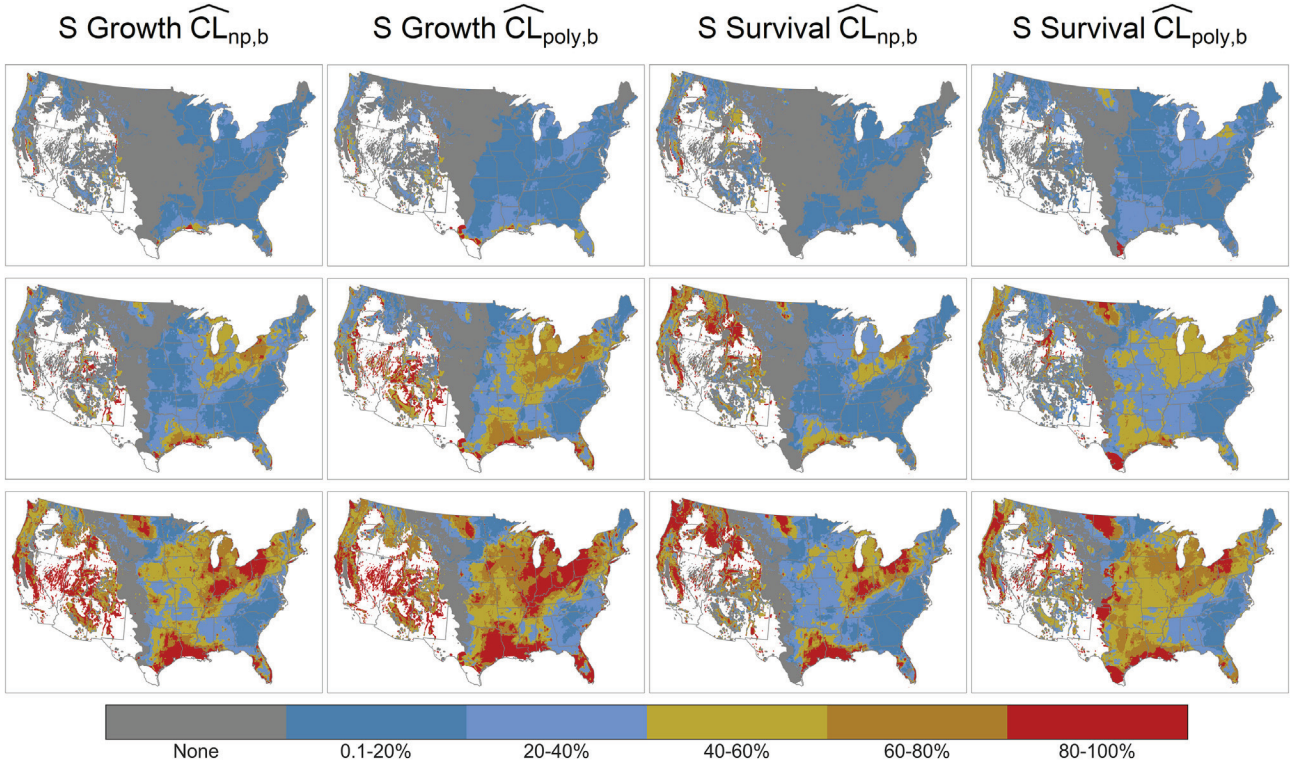



**Fig. 5.** Percentage of tree species exceeding their calculated S CL from $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ at high (upper 95th CI bound; top), median (median CL value; middle), and low (lower 95th CI bound; bottom) values within the bootstrapped CI. White indicates locations where none of the tree species for which a CL CI could be calculated are known to occur.

obtained between the methods. $\widehat{CL}_{poly,b}$ values consistently have lower percent of species exceeding the CL relative to $\widehat{CL}_{np,b}$ values across the conterminous United States for N deposition impacts on tree growth. For S deposition impacts on tree survival, the $\widehat{CL}_{poly,b}$ and $\widehat{CL}_{np,b}$ methods further indicate different spatial patterns in the proportion of tree species experiencing deposition exceeding the CL. The median $\widehat{CL}_{np,b}$ results in the highest proportion of exceedances in Indiana, Ohio, and eastern Texas, while the median $\widehat{CL}_{poly,b}$ results indicate a higher proportion of exceedances in the Dakotas and eastern Pennsylvania.

Maps comparing the CL CIs for each tree species with historical atmospheric deposition are provided in Datasets S1 and S2. The maps indicate the extent to which deposition levels fall within the CL CIs calculated for each species. For example, although paper birch $\widehat{CL}_{poly,b}$ CIs for tree growth impacts from atmospheric deposition of S cover a relatively small range of 2.16 kg S $ha^{-1}$ $yr^{-1}$, 77 % of the species range is impacted by deposition levels within the lower and upper bounds of the CI.

## 4. Discussion

Our analysis of effects from atmospheric N and S deposition on the growth and survival of 108 tree species demonstrates a novel application of ML to model the relationship between tree outcomes and atmospheric deposition and to constrain the uncertainty in that relationship. Substantial variability in the range of uncertainty of critical loads between species was observed in our model results, with some species showing large uncertainty ranges and others with very small uncertainty ranges. This variability can be attributed to differences in the range of deposition experienced by trees across species and variability between individual tree in their sensitivity to deposition. The results show that the range of uncertainty in CLs we calculated with this approach spans the actual atmospheric deposition that occurs over many species' geographic ranges. The range of the uncertainty is sufficiently large to warrant consideration in management and regulatory decision making.

The ML models we developed performed well in representing tree growth and survival in independent testing. When compared with previous work (Horn et al., 2018), the CLs established with the ML approach are comparable or often lower, and the model performance is improved by a factor of two. The improvement in the prediction of tree growth suggests that the CLs established using ML methods are likely to reflect the relationship between atmospheric deposition and tree growth with greater fidelity (Zhao and Hastie, 2021).

In addition to improved model performance, the ML methods used here provide several advantages. First, the ML techniques we used do not rely on parametric assumptions about the shape of the relationship between atmospheric deposition and tree growth and survival. In particular, the $\widehat{CL}_{stat}$ and $\widehat{CL}_{np,b}$ tests can establish CLs regardless of the shape of the dose response function (e.g., step function). Second, XGBoost represents variable interactions, such that our CLs account for the impact of environmental variables on tree sensitivity to a particular level of atmospheric deposition. Third, the use of PD to characterize the relationship between deposition and tree growth and survival, combined with bootstrapping, allows us to estimate the impact of N and S deposition while holding all other factors constant. This approach mitigates concerns related to confounding between N and S deposition. However, co-effects of N and S are also likely to occur. While these interactions are expected to be represented in the trained models, they are not captured by the PD approach use to quantify CLs for N and S individually. In future work, we plan to investigate these interactions directly. For these reasons, the bootstrap machine-learning ensemble approach we describe here is well suited to development of CLs and CL CIs for trees and likely could be applied effectively to other ecological endpoints.

Our methods quantified the uncertainty in CL calculations resulting from sampling error associated with the FIA tree database using a bootstrap approach. However, additional sources of uncertainty in CL values exist and were not assessed. Among these, uncertainty in the reported deposition may be an important source itself (Walker et al., 2019). Additional sources of uncertainty include potential temporal lags in ecosystem response, correlations among predictors, and spatiotemporal bias in FIA sampling schedules. These sources of uncertainty are common to all empirical CL methods that use similar sources of input data. Future work is needed to quantify the full range of uncertainty associated with CL estimates.

There are several opportunities to extend this work. The ML models we developed could be extended to incorporate additional environmental variables likely to impact tree growth and survival. In particular, soil characteristics and disturbances such as drought, short-term meteorological events, beetle infestation, or wildfire events could be incorporated. Future work could use model interpretation techniques (multivariate PD) to calculate how additional environmental factors influence tree sensitivity to atmospheric deposition. For example, such models could be used to investigate the relationship between soil pH and the sensitivity of trees to atmospheric deposition (Battles et al., 2014; Sullivan et al., 2013). Such relationships, once established, could be applied to develop spatially-explicit maps of CLs that vary as environmental conditions, besides climate, change across the landscape. This approach can be used to understand the impact of variability in the sensitivity of individual trees to atmospheric deposition on uncertainty in the species-level critical loads calculated here. Climate-related variables also present an opportunity to explore tree sensitivity under future climate change scenarios (Van Houtven et al., 2018). In addition, models that included ozone could be used to constrain the uncertainty of ozone critical levels, with implications for regulatory decision making.

## 5. Conclusions

The results we present here show that bootstrapped ensemble ML provides a robust approach to (1) estimating CLs for atmospheric nitrogen and sulfur deposition for the growth and survival of individual tree species across the coterminous U.S. and (2) constraining their uncertainty. These results can be used by land managers and regulators to guide decision making related to atmospheric deposition. For example, the U.S. EPA (U.S. Environmental Protection Agency, 2017) included CL information in a review of secondary standards for atmospheric nitrogen and sulfur oxides under the NAAQS.

## CRediT authorship contribution statement

**Nathan R. Pavlovic:** Conceptualization, Methodology, Writing – original draft, Supervision, Project administration, Funding acquisition. **Shih Ying Chang:** Methodology, Investigation, Visualization. **Jiaoyan Huang:** Validation, Data curation, Visualization. **Kenneth Craig:** Conceptualization, Writing – review & editing. **Christopher Clark:** Conceptualization, Writing – review & editing. **Kevin Horn:** Writing – review & editing, Data curation. **Charles T. Driscoll:** Conceptualization, Writing – review & editing.

## Data availability

Data will be made available on request.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We gratefully acknowledge Alejandro Schuler, Jennifer Phelan, and Michael Bell for valuable discussion to support this analysis. Funding: This work was supported by the Electric Power Research Institute. The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of the U.S. EPA or any other federal agency.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.159252.

## References


Battles, J.J., Fahey, T.J., Driscoll, C.T., Blum, J.D., Johnson, C.E., 2014. Restoring soil calcium reverses forest decline. Environ. Sci. Technol. Lett. 1 (1), 15–19. https://doi.org/10.1021/ez400033d Available at.

Bekkerman, R., 2015. The present and the future of the KDD cup competition: an outsider's perspective. Available at https://www.linkedin.com/pulse/present-future-kdd-cup-competition-outsiders-ron-bekkerman.

Blett, T.F., Lynch, J.A., Pardo, L.H., Huber, C., Haeuber, R., Pouyat, R., 2014. FOCUS: a pilot study for national-scale critical loads development in the United States. Sci. Policy 38, 225–262.

Burns, D.A., Sullivan, T.J., 2015. **USE 17150**Critical Loads of Atmospheric Deposition to Adirondack Lake Watersheds: A Guide for Policymakers. Prepared for the New York State Energy Research and Development Authority, Albany, NY.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. Paper From the Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Conference, San Francisco, CA, August. ACM, pp. 785–794 https://doi.org/10.1145/2939672.2939785 Available at http://arxiv.org/abs/1603.02754%0Ahttp://dx.doi.org/10.1145/2939672.2939785.

Clark, C.M., Simkin, S.M., Allen, E.B., Bowman, W.D., Belnap, J., Brooks, M.L., Collins, S.L., Geiser, L.H., Gilliam, F.S., Jovan, S.E., Pardo, L.H., Schulz, B.K., Stevens, C.J., Suding, K.N., Throop, H.L., Waller, D.M., 2019. Potential vulnerability of 348 herbaceous species to atmospheric deposition of nitrogen and sulfur in the United States. Nat. Plants 5, 697–705 July 2019.

CLRTAP, 2004. Manual on Methodologies and Criteria for Modelling and Mapping Critical Loads and Levels and Air Pollution Effects, Risks and Trends. Available atUmwelbundesamt, Berlin. https://www.umweltbundesamt.de/en/cce-manual.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874.

Fenn, M.E., Lambert, K.F., Blett, T.F., Burns, D.A., Pardo, L.H., Lovett, G.M., Haeuber, R.A., Evers, D.C., Driscoll, C.T., Jeffries, D.S., 2011. Setting limits: Using air pollution thresholds to protect and restore US ecosystems. Issues Ecol. (14).

Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.

Geiser, L.H., Jovan, S.E., Glavich, D.A., Porter, M.K., 2010. Lichen-based critical loads for atmospheric nitrogen deposition in Western Oregon and Washington Forests, USA. Environ. Pollut. 158, 2412–2421. https://doi.org/10.1016/j.envpol.2010.04.001.

Geiser, L.H., Nelson, P.R., Jovan, S.E., Root, H.T., Clark, C.M., 2019. Assessing ecological risks from atmospheric deposition of nitrogen and sulfur to U.S. forests using epiphytic macrolichens. Diversity 2019 (11), 87. https://doi.org/10.3390/d11060087.

Horn, K.J., Thomas, R.Q., Clark, C.M., Pardo, L.H., Fenn, M.E., Lawrence, G.B., Perakis, S.S., Smithwick, E.A.H., Baldwin, D., Braun, S., Nordin, A., Perry, C.H., Phelan, J.N., Schaberg, P.G., Clair, S.B.S., Warby, R., Watmough, S., 2018. Growth and survival relationships of 71 tree species with nitrogen and sulfur deposition across the conterminous U.S. Plos One 13 (10), e0205296. https://doi.org/10.1371/journal.pone.0205296.

Kobe, R.K., 1996. Intraspecific variation in sapling mortality and growth predicts geographic variation in forest composition. Ecol. Monogr. 66 (2), 181–201.

Kvålseth, T.O., 1983. Cautionary note about R 2 Am. Stat. 39 (4), 279–285. Available at:. https://doi.org/10.1080/00031305.1985.10479448 (10.1080/00031305.1985.10479448), July 1.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7 (1), 3–10.

Little, E.L., 1971. Atlas of United States Trees, Volume 1, Conifers and Important Hardwoods. 1. U.S. Department of Agriculture Miscellaneous Publication 1146 , p. 200. https://doi.org/10.5962/bhl.title.130546. Available at https://www.biodiversitylibrary.org/bibliography/130546.

Little, E.L., 1976. Atlas of United States Trees, Volume 3, Minor Western Hardwoods. 3. U.S. Department of Agriculture Miscellaneous Publication 1314, p. 290. https://doi.org/10.5962/bhl.title.65782 Available at doi:10.5962/bhl.title.65782.

Little, E.L., 1977. Atlas of United States Trees, Volume 4, Minor Eastern Hardwoods. 4. U.S. Department of Agriculture Miscellaneous Publication 1342, p. 230. https://doi.org/10.5962/bhl.title.65458 Available at doi:10.5962/bhl.title.65458.

Little, E.L., 1978. Atlas of United States Trees, Volume 5, Florida. 5. U.S. Department of Agriculture Miscellaneous Publication 1361, p. 262. https://doi.org/10.5962/bhl.title.65462 Available at doi:10.5962/bhl.title.65462.

Ma, J., Yu, Z., Qu, Y., Xu, J., Cao, Y., 2019. Application of the XGBoost Machine Learning method in PM2.5 prediction: a case study of Shanghai. Aerosol Air Qual. Res. 20. https://doi.org/10.4209/aaqr.2019.08.0408 January.

McNulty, S.G., Cohen, E.C., Myers, J.A.M., Sullivan, T.J., Li, H., 2007. Estimates of critical acid loads and exceedances for forest soils across the conterminous United States. Environ. Pollut. 149 (3), 281–292.

Molnar, C., 2019. Interpretable machine learning: a guide for making black box models explainable Available at https://christophm.github.io/interpretable-ml-book/ April 8, 2021.

Nilsson, J., Grennfelt, P., 1988. Critical loads for sulphur and nitrogen. Air Pollut.Ecosyst. 85–91.

Pardo, L.H., Fenn, M., Goodale, C.L., Geiser, L.H., Driscoll, C.T., Allen, E., Baron, J., Bobbink, R., Bowman, W.D., Clark, C., Emmett, B., Gilliam, F.S., Greaver, T., Hall, S.J., Lilleskov, E.A., Liu, L., Lynch, J., Nadelhoffer, K., Perakis, S., Robin-Abbott, M.J., Stoddard, J., Weathers, K., Dennis, R.L., 2011. Effects of nitrogen deposition and empirical critical loads for nitrogen for ecological regions of the United States. Ecol. Appl. 21 (8), 3049–3082.

Pardo, L.H., Coombs, J.A., Robin-Abbott, M.J., Pontius, J.H., D'Amato, A.W., 2019. Tree species at risk from nitrogen deposition in the northeastern United States: a geospatial analysis of effects of multiple stressors using exceedance of critical loads. For. Ecol. Manag. 454, 117528.

Phelan, J.N., Cummings, T.G., Bell, M.D., Lynch, J.A., 2018. Critical Load – A Tool to Evaluate the Sensitivity of Terrestrial and Aquatic Ecosystems to Air Pollution (version 1.0).

Program, N.A.D., 2021. National trends network. Available at https://nadp.slh.wisc.edu/.

Schwede, D.B., Lear, G.G., 2014. A novel hybrid approach for estimating total deposition in the United States. Available atAtmos. Environ. 92, 207–220. https://doi.org/10.1016/j.atmosenv.2014.04.008. http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1219&context=usepapapers.

Simkin, S.M., Allen, E.B., Bowman, W.D., Clark, C.M., Belnap, J., Brooks, M.L., Cade, B.S., Collins, S.L., Geiser, L.H., Gilliam, F.S., Jovan, S.E., Pardo, L.H., Schulz, B.K., Stevens, C.J., Suding, K.N., Throop, H.L., Waller, D.M., 2016. Conditional vulnerability of plant diversity to atmospheric nitrogen deposition across the United States. Proc. Natl. Acad. Sci. 113, 4086–4091.

Sullivan, T.J., Lawrence, G.B., Bailey, S.W., Beier, C.M., Weathers, K.C., Bishop, D.A., McDonnell, T.C., McPherson, G.T., 2013. Effects of acidic deposition and soil acidification on sugar maple trees in the Adirondack Mountains, New York. Environ. Sci. Technol. 47 (22), 12687–12694. https://doi.org/10.1021/es401864w Available at.

U.S. Environmental Protection Agency, 2011. Policy assessment for the review of the secondary National Ambient Air Quality Standards for oxides of nitrogen and oxides of sulfur. EPA-452/R-11-005a. Available at http://www.epa.gov/ttnnaaqs/standards/no2so2sec/data/20110204pamain.pdf.

U.S. Environmental Protection Agency, 2017. Integrated Review Plan for the Secondary National Ambient Air Quality Standards for Ecological Effects of Oxides of Nitrogen, Oxides of Sulfur and Particulate Matter. EPA-452/R-17-002. Available at April 8, 2021.

Van Houtven, G., Phelan, J., Clark, C., Sabo, R.D., Buckley, J., Thomas, R.Q., Horn, K., LeDuc, S.D., 2018. Nitrogen deposition and climate change effects on tree species composition and ecosystem services for a forest cohort. Ecol. Monogr. 89 (2). https://doi.org/10.1002/ecm.1345 December 10.

Van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fulé, P.Z., Harmon, M.E., Larson, A.J., Smith, J.M., Taylor, A.H., Veblen, T.T., 2009. Widespread increase of tree mortality rates in the western United States. Science 323 (5913), 521–524.

Walker, J.T., Bell, M.D., Schwede, D., Cole, A., Beachley, G., Lear, G., Wu, Z., 2019. Aspects of uncertainty in total reactive nitrogen deposition estimates for North American critical load applications. Sci. Total Environ. 690, 1005–1018.

Wu, W., Driscoll, C.T., 2010. Impact of climate change on three-dimensional dynamic critical load functions. Environ. Sci. Technol. 44 (2), 720–726.

Zhao, Q., Hastie, T., 2021. Causal interpretations of black-box models. J. Bus. Econ. Stat. 39 (1), 272–281. https://doi.org/10.1080/07350015.2019.1624293.