forests

MDPI

Article

Evaluating Inter-Rater Reliability and Statistical Power of Vegetation Measures Assessing Deer Impact

Danielle R. Begley-Miller [1,*], Duane R. Diefenbach [2], Marc E. McDill [3], Christopher S. Rosenberry [4] and Emily H. Just [5]

1 Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, University Park, PA 16802, USA
2 U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University, University Park, PA 16802, USA; drd11@psu.edu
3 Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, USA; mem14@psu.edu
4 Pennsylvania Game Commission, Harrisburg, PA 17110, USA; chrosenber@pa.gov
5 Department of Conservation and Natural Resources, Bureau of Forestry, Harrisburg, PA 17105, USA; emjust@pa.gov
* Correspondence: dbegley-miller@teatown.org; Tel.: +1-914-7622-912 (ext. 119)

Received: 24 August 2018; Accepted: 15 October 2018; Published: 25 October 2018



Abstract: Long-term vegetation monitoring projects are often used to evaluate how plant communities change through time in response to some external influence. Here, we evaluate the efficacy of vegetation monitoring to consistently detect changes in white-tailed deer browsing effects. Specifically, we compared inter-rater reliability (Cohen's κ and Lin's concordance correlation coefficient) between two identically trained field crews for several plant metrics used by Pennsylvania state agencies to monitor deer browsing impact. Additionally, we conducted a power analysis to determine the effect of sampling scale (1/2500th or 1/750th ha plots) on the ability to detect changes in tree seedling stem counts over time. Inter-rater reliability across sampling crews was substantial for most metrics based on direct measurements, while the observational based Deer Impact Index (DII) had only moderate inter-rater reliability. The smaller, 1/2500th ha sampling scale resulted in higher statistical power to detect changes in tree seedling stem counts due to reduced observer error. Overall, this study indicates that extensive training on plant identification, project protocols, and consistent data collection methods can result in reliable vegetation metrics useful for tracking understory responses to white-tailed deer browsing. Smaller sampling scales and objective plant measures (i.e., seedling counts, species richness) improve inter-rater reliability over subjective measures of deer impact (i.e., DII). However, considering objective plant measures when making a subjective assessment regarding deer browsing effects may also improve DII inter-rater reliability.

Keywords: vegetation monitoring; observer error; statistical power; white-tailed deer; inter-rater reliability

1. Introduction

Long-term ecological monitoring is crucial for tracking how forest ecosystems change through time, especially in the context of human influence [1]. The issue with long-term monitoring, however, is that there is inherent year-to-year variability associated with unmeasured conditions (i.e., noise [2]), and sampling variability (i.e., sampling error) contributes to imperfect collection methods [3,4]. Long-term ecological data requires sufficient statistical power to detect change in measured conditions following some external influence, whether it be climate [5], invasion [6], disturbance [7], pest/disease [8], or

some other environmental variable [9]. Statistical power is defined as the inverse of the probability of a type II error, or, more simply, one minus the probability of failing to reject the null hypothesis when the null hypothesis is false (a "false negative"). The effectiveness of long-term monitoring programs is tied directly to specific, achievable management objectives partnered with an appropriate sampling design to address those objectives [10].

Forest inventories are commonly used by land managers to document baseline forest conditions, and, when resurveyed, evaluate how forest vegetation changes through time [11–15]. Nationally, the United States Department of Agriculture (USDA) Forest Service inventories permanent plots across each state annually, visiting 20% of all plots per state per year [16]. Prior to 1998, data collection protocols for the Forest Service's Forest Inventory and Analysis (FIA) program were not nationally consistent, resulting in variable trends by state across several forest measures [17]. Many studies have shown that inconsistent data collection, observer bias, and varying sampling methods have an effect on data analysis and interpretation [3,10,17,18]. To combat these issues and improve statistical power, researchers recommend extensive training on data collection protocols, specific objectives, consistent observers, and/or quality control data assessments [18].

Over time, the Forest Service has enhanced the FIA program and expanded data collection to include soils, understory vegetation, white-tailed deer browsing impact, and other limiting factors on a subset of plots per state [19]. The objective of the FIA program is to monitor changes in abundance and species composition of all US forests, and correlate those changes with other measures of ecosystem health [20]. The Pennsylvania Department of Conservation and Natural Resources, Bureau of Forestry adopted similar enhanced forest inventory protocols (Continuous Forest Inventory, or CFI) in 1997 to continuously monitor 890,000 ha of public state forest land [21]. Both agencies have similar missions, and work to promote the long-term viability and productivity of forested lands for future generations [22,23]. In addition to sustainable forest management, the Bureau of Forestry is responsible for the protection of rare and endangered plants, and its inventory protocols more intensively sample understory vegetation at all CFI plot locations compared to FIA methods [21]. The FIA protocol, however, provides more data on overstory conditions, stocking rates, and available timber [24].

In 2014, the Bureau of Forestry increased the number of vegetation monitoring plots and implemented a new monitoring program (the Vegetation Impact Protocol, or VIP) specifically designed to assess the effect of deer on understory plant communities [25,26]. White-tailed deer browsing has been identified as an important driver of plant community composition throughout the white-tailed deer range [13,27–29], and selective browsing by white-tailed deer has the ability to reduce understory diversity and structure in areas where deer populations are near carrying capacity [30–32]. The VIP is modeled after CFI protocols, but VIP primarily monitors understory phyto-indicators of deer browsing impact. Ultimately, the Bureau of Forestry uses VIP data (tree seedling stem counts and the abundance/reproductive status of browse-sensitive herbaceous plants) to decide whether to enroll an area of state forest land into the deer management assistance program (DMAP). Following a DMAP permit application process and the appropriate justification by the Bureau of Forestry, the Pennsylvania Game Commission (PGC) issues additional antlerless tags for purchase by the hunting community in these select areas on public, state forest land. The Bureau of Forestry continuously monitors DMAP areas to determine when enrollment status should change due to a change in understory vegetation.

The PGC also evaluates deer effects on forests to meet agency objectives, which includes maintaining deer impact that supports sustainable forests, and implementing measures of deer impact in forests to improve effectiveness of deer management programs [33]. The PGC uses FIA habitat data, including the FIA deer impact index (DII), from all private and public forested lands in Pennsylvania to make deer management decisions and set tag allocations for the upcoming hunting season for each Wildlife Management Unit (WMU) across the state [33]. The integration of habitat data and deer impact assessments into both federal and state agency decision models highlights the need for consistent evaluations of habitat characteristics across different sampling scales and vegetation measures.

We developed a forest inventory approach that merges FIA and VIP data protocols to intensively monitor both overstory and understory vegetation data and evaluate both FIA and VIP sampling designs. Because the Bureau of Forestry collects understory inventory data on 5 1/2500th ha subplots per sampling site, and the FIA protocol collects understory data on 4 larger, 1/750th ha subplots per site, we chose to collect data at both scales at 5 subplots for each sampling location. Our goal was to have data that were comparable to both FIA and VIP data, specifically focusing on the effect of scale on observer error and data consistency. Larger search areas may result in increased sampling errors as the amount of vegetation encountered increases [34]. We also assessed deer impact using the FIA DII, a subjective measure based on observer assessments of vegetation conditions and browsing levels at a site. In general, subjective measures like the DII may be less reliable than vegetation metrics based on objectively collected data.

We compared inter-rater reliability of both categorical and continuous vegetation metrics between two identically-trained field crews separately for each scale of data collection, and expected reliability to be inversely correlated to rates of sampling (observer) error. We hypothesized that if reliability is inversely related to error, there would be greater inter-rater reliability at smaller sampling scales than larger ones (H1). For categorical variables, we specifically compared the DII, the total number of subplots at a sampling site with tree regeneration present, the total number of subplots at a sampling site with the presence of lilaceous forest herbs preferred by white-tailed deer (phyto-indicators of deer browsing effects, or indicators), and indicator species' richness between crews with Cohen's Kappa (κ). For continuous variables, we compared the total counts of all tree seedling taxa <2.54 cm diameter at breast height (DBH), total counts of all tree seedling taxa >0.3 m tall and <2.54 cm DBH, total counts of all tree seedlings paired by taxon, and total counts of all tree seedlings >0.3 m tall and <2.54 cm DBH paired by taxon at each sampling location.

To further assess the effects of sampling error on detecting changes in vegetation characteristics, we conducted a power analysis across a range of sample sizes using tree seedling data simulated from real-world parameters for each scale. Forested ecosystems are ecologically complex [2,35], and studies in forest systems often require large sample sizes ($n > 30$) or very specific sampling methods to detect measurable changes in forest conditions [36]. Increased error rates from inaccurate counts of tree seedlings reduces the statistical power to detect changes in deer browsing pressure when these counts are used to assess deer browsing effects on tree regeneration [31,37–40]. To evaluate the effect of count errors on statistical power to detect changes in tree seedling abundance, we simulated both initial starting tree seedling counts (time 1 data) and a mean stem change to those counts (time 2 data) using means and variances derived from actual field-collected and CFI data. We added sampling error to each data set based on tree seedling counts for both crews at overlapping sampling locations using the residuals of a generalized linear model. Then, we compared whether we could detect a difference in mean stem counts between both data sets (time 1 vs. time 2) for a given sample size. Power analysis specifics are described in detail below in the Materials and Methods section and illustrated in Figure 1. We hypothesized that if sampling variance was related to sampling scale, the smaller scale of data collection would have increased statistical power due to reduced variability in stem counts between crews (H2).

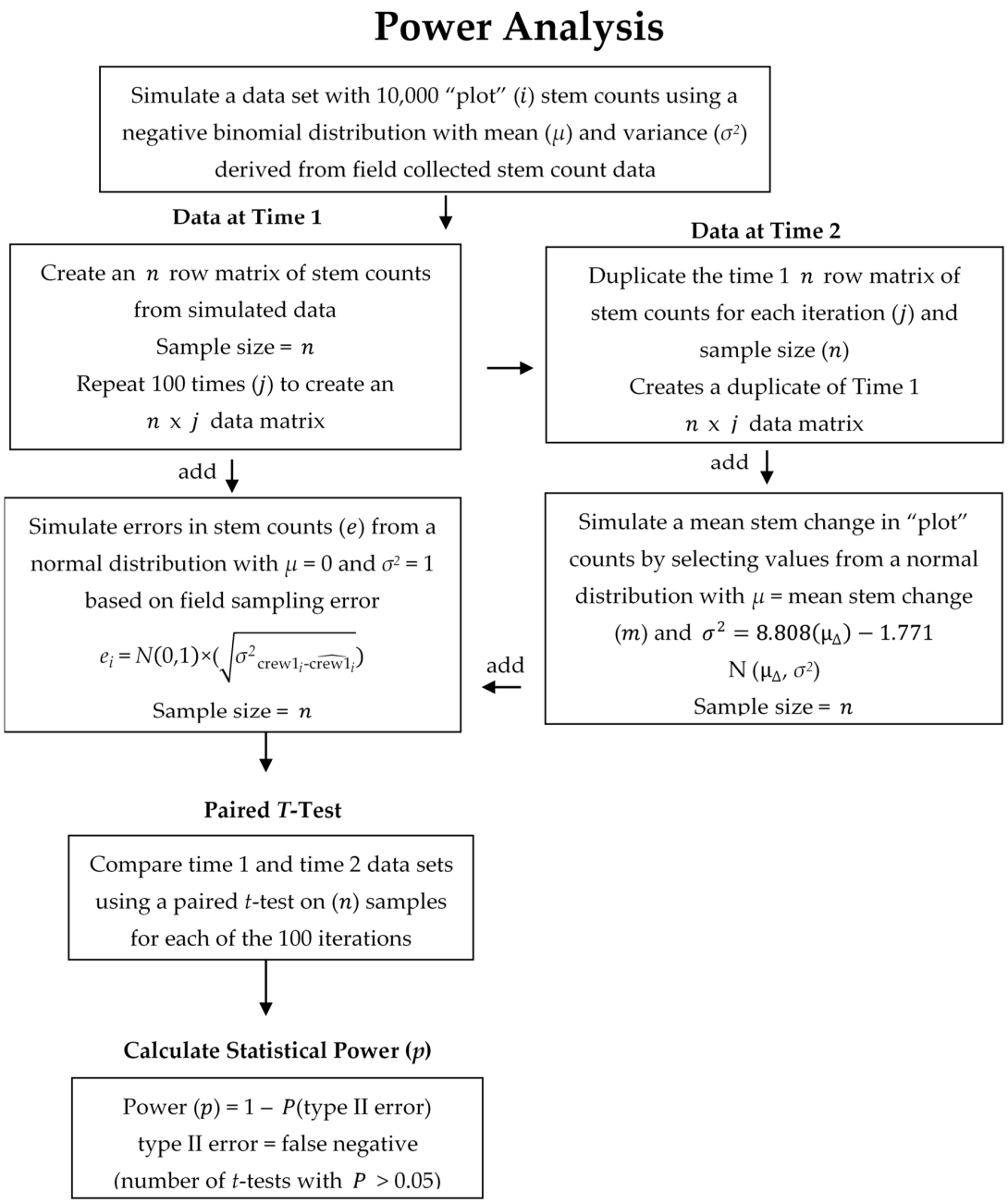


Figure 1. Illustration of all power analysis steps including simulations of each sample size (n = 20, 30, 40, 50, 60, 70, 80, 90, 100) and mean stem change (m = 0, 0.5, 1.0, 1.5, 2.0) scenario, beginning with data simulation and proceeding to all other steps in order following diagram arrows. All sample size and mean stem change scenarios were replicated 100 times to calculate a mean level of power for each group.

2. Materials and Methods

Three separate field crews collected vegetation data from 26 May to 10 August 2015 as part of a larger study assessing the effects of white-tailed deer on Pennsylvania forest plant communities. We trained all crews together from 11 to 22 May 2015 on plant identification, data collection protocols, and assessment of white-tailed deer browsing impact on plot understory vegetation. Following training, crews independently collected data at different plot locations across three state forests: The Susquehannock State forest in the Appalachian Plateau Province of north-central Pennsylvania, and the Rothrock and Bald Eagle State forests located in the Ridge and Valley Physiographic Province of

central Pennsylvania. In total, each crew visited either 68, 24, or 59 plot locations over the summer sampling season, depending on their crew assignment. Two of those crews in the Rothrock and Bald Eagle State Forests collected data at 13 overlapping sampling sites independently so that we could assess inter-rater reliability. Sampling locations for all 3 crews are shown in Figure 2.

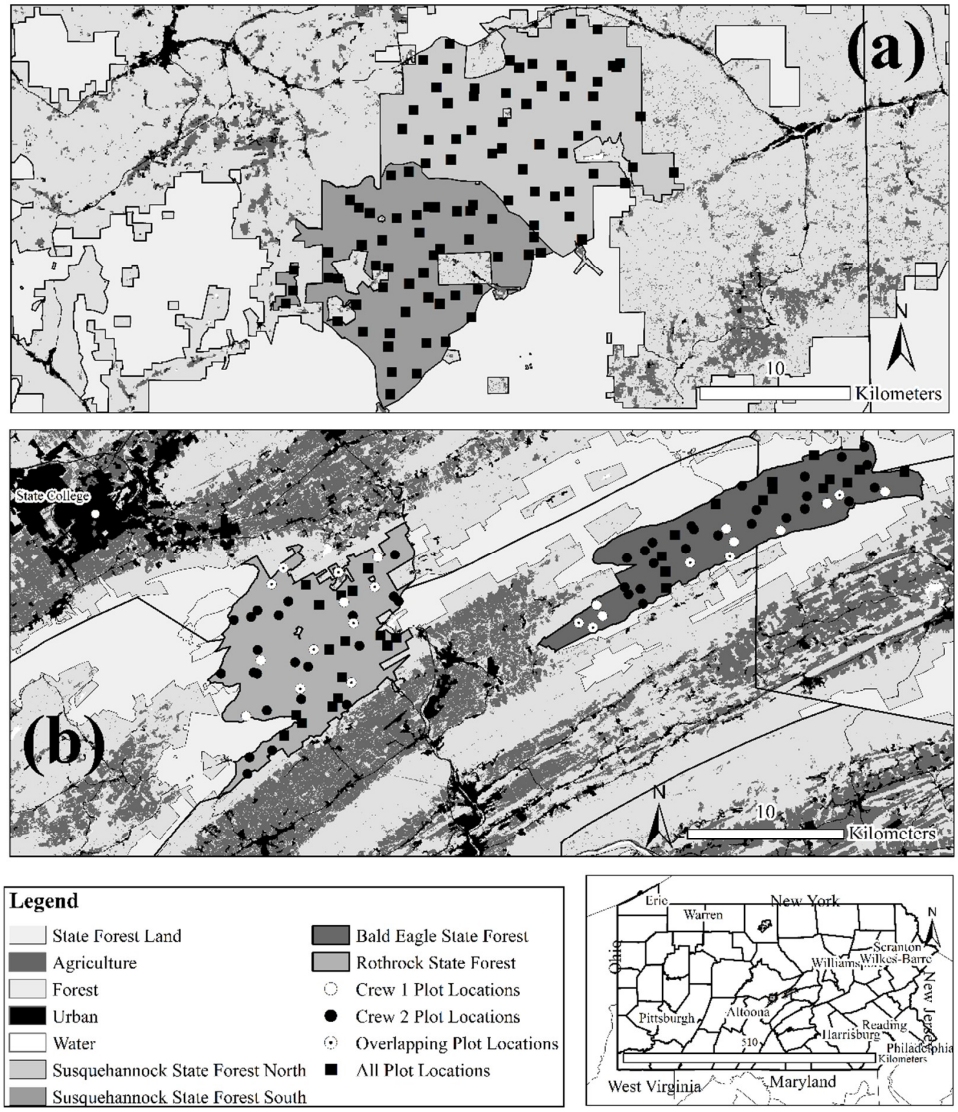


Figure 2. Map of the study area and Pennsylvania land cover (agriculture, forest, urban, and water). The study areas are highlighted by dark gray polygons (Susequhannock State Forest South, (**a**) bottom; Susquehannock State Forest North, (**a**) top; Rothrock State Forest, (**b**) left; Bald Eagle State Forest, (**b**) right), and are located on state forest land (light polygons). White circles represent plots sampled by crew 1 only (11), black circles represent plots sampled by crew 2 only (46), and white circles with black dots represent plots sampled independently by both crews (13). Black triangles represent additional plots sampled in the northern study area in 2015 (68).

Nested within each plot location was a network of 5 subplots oriented such that a single subplot's center was not more than 36.5 m straight-line distance from the next nearest subplot's center (Figure 3). Further nested within each subplot was a smaller survey area, or sampling zone, where crews conducted full vegetation inventories at two scales (1/2500th ha and 1/750th ha). We defined subplots as the 7.32 m radial area around a center stake in the middle of the subplot, and sampling zones as the 2.07 m radial area around a stake located 3.66 m due east (90° azimuth) of subplot center. Within a sampling zone, a smaller 1/2500th ha microplot is nested within a larger 1/750th ha microplot. The

larger, 1/750th ha microplot replicates Forest Inventory and Analysis (FIA) sampling protocols [24], and the smaller, 1/2500th ha microplot replicates the Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry's sampling protocol to assess deer impact [26]. Crews collected vegetation data independently for each scale at all 5 subplots at each sampling location.

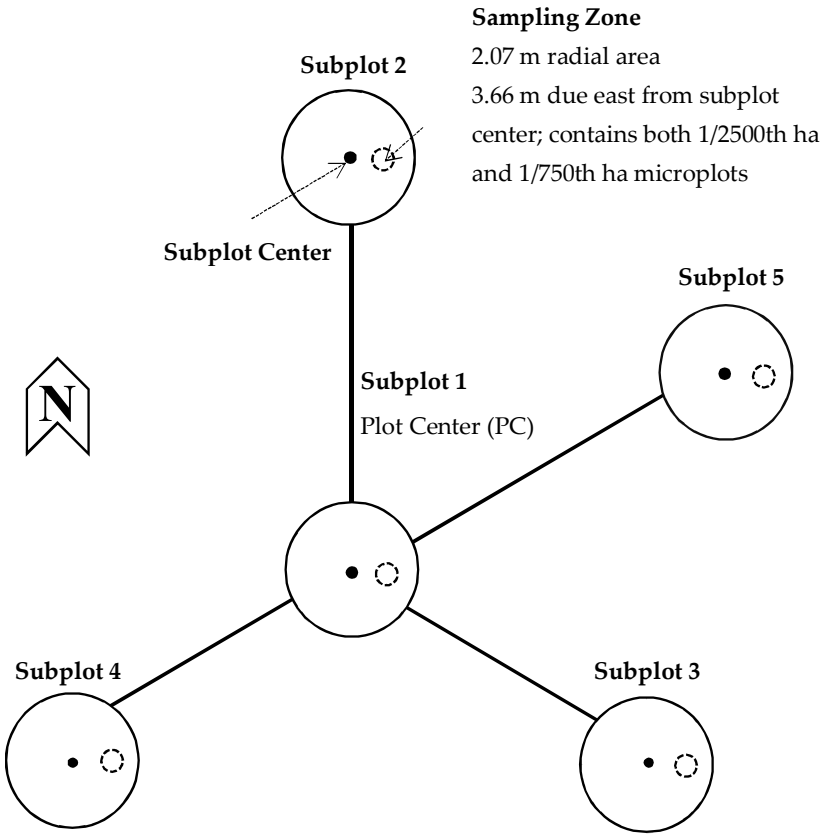


Figure 3. Configuration of subplots 1–5 at each plot location. We defined the center of subplot 1 as plot center (PC), and the area inside each subplot circle represents 1/60th ha (1/24th acre; 168 m^2). Sampling zone circles represent a 1/750th ha microplot (small dashed circle; 1/300th acre; 13.5 m^2), and nested within each sampling zone is a second 1/2500th ha microplot (1/1000th acre; 4.05 m^2; not illustrated). The larger, 1/750th ha microplot replicates Forest Inventory and Analysis (FIA) sampling protocols, whereas the smaller, 1/2500th ha microplot replicates the Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry's sampling protocols.

2.1. Study Area Characteristics

The study area was primarily forested with even-aged (75–100-year-old) oak-hickory hardwood stands, and located within the Ridge and Valley Physiographic Province of Pennsylvania. On average, the growing season for the region was 182 days from 22 April to 21 October [41]. The climate was temperate with 104 cm of mean annual precipitation, and mean summer temperatures range from 16 °C at night to 29 °C during the day [42]. Across sampling locations, the overstory was comprised mainly of oak, including red (*Quercus rubra* L.) and chestnut oak (*Quercus montana* Willd.), and the understory was dominated by ericaceous shrubs, including mountain laurel (*Kalmia latifolia* L.), huckleberry (*Gaylussacia* spp.), and blueberry (*Vaccinium* spp.). Elevation of all plots ranged from 400 to 700 m above sea level.

2.2. Vegetation Monitoring

From 26 May to 10 August 2015, crews conducted 151 vegetation inventories at 138 plot locations (crew 1 = 24; crew 2 = 59; crew 3 = 68). Crew 1 and crew 2 visited 13 of those locations independently

within the same 10-day window (mean number of days between crew visits was 3 days, with a range of 0 to 10 days). Crews identified all understory taxa at each microplot scale, and counted total numbers of tree seedlings (any live, arborescent, woody species <2.54 cm DBH) by taxon. Crews also counted the number of ramets of phyto-indicators of deer browsing effects (Indian cucumber-root (*Medeola virginiana* L.), Canada mayflower (*Maianthemum canadense* Desf.), Trillium (*Trillium* spp.), false Solomon's seal (*Maianthemum racemosum* (L.) Link), and true Solomon's seal (*Polygonatum* spp.)) by taxon [43–45]. Lastly, crews assessed deer browsing impact at a site based on FIA deer impact assessment criteria (Table 1).

Table 1. Forest Inventory and Analysis (FIA) deer impact index assessment criteria.

Code	Definition
1	Very Low—Plot is inside a well-maintained deer exclosure.
2	Low—No browsing observed, vigorous seedlings present (no deer exclosure present).
3	Medium—Browsing evidence observed but not common, seedlings present.
4	High—Browsing evidence common OR seedlings are rare.
5	Very High—Browsing evidence omnipresent OR forest floor bare, severe browse line.

2.3. Data Analysis

We assessed the level of inter-rater reliability between the two Rothrock and Bald Eagle survey crews across several plant metrics at both microplot scales to assess consistency in data collection. We also compared plot means of tree seedling counts >0.3 m tall and <2.54 cm diameter between the two crews to determine count sampling error, and conducted a power analysis to determine what sample size would be necessary to detect a change in mean stem counts given observer error. We evaluated the two crew's agreement for both categorical and continuous vegetation survey data that would help assess the effects of deer browsing on tree seedlings and herbaceous plants at both microplot scales. Categorical data assessed included: (1) number of microplots per plot with at least 1 (1/2500th ha scale) or 4 (1/750th ha scale) tree seedling(s) >0.3 m tall and <2.54 cm diameter (see Table A1 for full list of taxa, of which nearly all were preferred browse by white-tailed deer), (2) the number of microplots per plot with at least 1 phyto-indicator of deer browsing, (3) indicator species richness across all microplots, and (4) deer impact assessment at the plot level. Continuous variables compared were: (5) count of all tree seedlings per plot, (6) count of all tree seedlings >0.3 m tall and <2.54 cm diameter per plot, (7) count of all seedlings by taxon per plot, and (8) count of all seedlings >0.3 m tall and <2.54 cm diameter by taxon per plot.

Inter-rater reliability of categorical data was assessed using Cohen's Kappa (2 raters) in the "irr" package (version 0.84, Hamburg, Germany) [46] in R (version 3.5.1, Vienna, Austria) [47]. Cohen's Kappa (κ) is a statistical measure of agreement between two independent raters that accounts for potential agreement by chance alone [48,49]. Positive values of κ indicate a level agreement beyond what would be expected randomly, but interpreting exact values is subjective. Landis and Koch (1977) developed 5 categories of agreement for Cohen's κ that have been consistently reported in the medical literature: $\kappa < 0.20$ is poor, $\kappa = 0.20$–0.40 is fair, $\kappa = 0.41$–0.60 is moderate, $\kappa = 0.61$–0.80 is good, and $\kappa > 0.80$ is very good. We use these categories in our interpretation of Cohen's κ.

For continuous variables, we assessed rater agreement using Lin's concordance correlation coefficient (CCC) in the "cccrm" package (version 1.2.1, Barcelona, Spain) [50] in R (version 3.5.1, Vienna, Austria). The concordance correlation coefficient is a measure of the level of deviation from an expected 1 to 1 ratio of measurement agreement between raters [51]. Like with Cohen's κ, CCC gets its origin from the medical field and interpretation of CCC values is subjective. McBride (2005) proposed 4 criteria for agreement to ensure consistency in reporting: CCC < 0.90 is poor, CCC = 0.90–0.95 is moderate, CCC = 0.96–0.99 is substantial, and CCC > 0.99 is almost perfect. We use these categories to present results from CCC.

Following assessments of inter-rater reliability, we conducted a power analysis to determine the sample size required to detect varying levels of change in mean tree seedling stems given observer error. We simulated a single data set of stem counts comprised of 10,000 "plots." Mean (μ) and variance (σ^2) of simulated data was set equal to the μ and σ^2 of stem counts from all plots sampled in 2015 (Figure 2). The data set was generated using a negative binomial distribution for stem counts (y) for each plot (i):

$$y_i = NB(r,\ p) \tag{1}$$

where

$$r = (\mu \times \mu^2)/\sigma^2 \tag{2}$$

and

$$p = r/(r + \mu) \tag{3}$$

The first data set, the "time 1" data set, was created by randomly sampling n "plots" with replacement from the simulated data for each sample size (n). Sample sizes ranged from 20 to 100 in increments of 10, and simulated data were sampled 100 times to create a data matrix with n rows and 100 columns. We simulated the second data set, a the "time 2" data set, by duplicating the "time 1" data set and adding in a mean change in stem counts (m) simulated from a normal distribution with a constant mean (u_Δ) and mean-dependent variance (σ^2). The mean–variance relationship was modeled directly from past CFI data (see Figure A1 for details on the mean–variance relationship):

$$m_i \ \sim \ N(\mu_\Delta, \sigma^2) \tag{4}$$

where

$$\sigma^2 = \ 8.808(\mu_\Delta) - 1.771 \tag{5}$$

Lastly, we added observer error to the "time 2" data set using the residuals from a negative binomial regression of crew 1 and crew 2 mean stem counts for each of the 13 plot locations (i) sampled by both crews. Crew 1 stem counts were considered the baseline "no error" values in the regression because both crew members were more experienced in both plant identification and project protocols (both individuals were returning members from the 2014 sampling season). Any differences in counts from crew 1 to crew 2 was attributed to crew 2 observer error using the following equation:

$$crew\ 2_i = \ crew\ 1_i + \ \varepsilon_i \tag{6}$$

where

$$\varepsilon_i \ \sim \ normal(0,\ \sigma^2) \tag{7}$$

and error values for each simulation were calculated as:

$$error_i = N(0,1) \times (\sqrt{\sigma^2{}_{crew1_i - \hat{crew1}_i}}) \tag{8}$$

where $\hat{crew1}_i$ is the predicted stem count value for $crew1_i$ given the linear relationship between crew stem counts. For each of the 100 simulations, error values were rounded to the nearest whole number, and added to the "error" dataset to represent expected observer error. Any values less than 0 were truncated at 0.

We compared output for each of the 100 simulations for both the "time 1" and "time 2" data sets using a paired t-test to determine the probability of committing a type II error (false negative). Because both datasets were initially simulated using the same values and the "time 2" data set had a known level of mean stem change applied, the failure to detect a difference mean stem counts within the 100 simulations ($p > 0.05$) was solely attributed to simulated observer error. We repeated this process for 5 simulated changes in mean stem counts per plot ranging from 0.5 to 2 in 0.5 increments for each

different sample size (n). We calculated the power (p) to detect change as 1—probability of a type II error (total number of tests with $p > 0.05$) for each mean stem change and sample size combination. We repeated the entire sampling process, including generating "time 1" and "time 2" data, 100 times and report the mean level of power across all simulations. These power analysis steps are illustrated in Figure 1.

3. Results

Inter-rater reliability (IRR) among categorical variables was lowest for the DII (κ = 0.54), but agreement was perfect for both the number of microplots with indicator plants (κ = 1.0) and indicator richness (κ = 1.0; Figure 4a) at both sampling scales. There was no difference in IRR across scale for either indicator plant measure, but for the number of microplots with tree regeneration, IRR was higher at the 1/2500th ha scale (κ = 0.89) than the 1/750th ha scale (κ = 0.72). There was no difference in agreement classification for total counts of all tree seedlings across scale (Figure 4b); agreement was substantial for both the 1/2500th ha (CCC = 0.96) and 1/750th ha (CCC = 0.98). There were discrepancies in agreement category across the 3 remaining tree seedling counts across scale (1/2500th = substantial, 1/750th ha = near perfect), but overlapping confidence intervals indicate that IRR differences were not statistically significant.

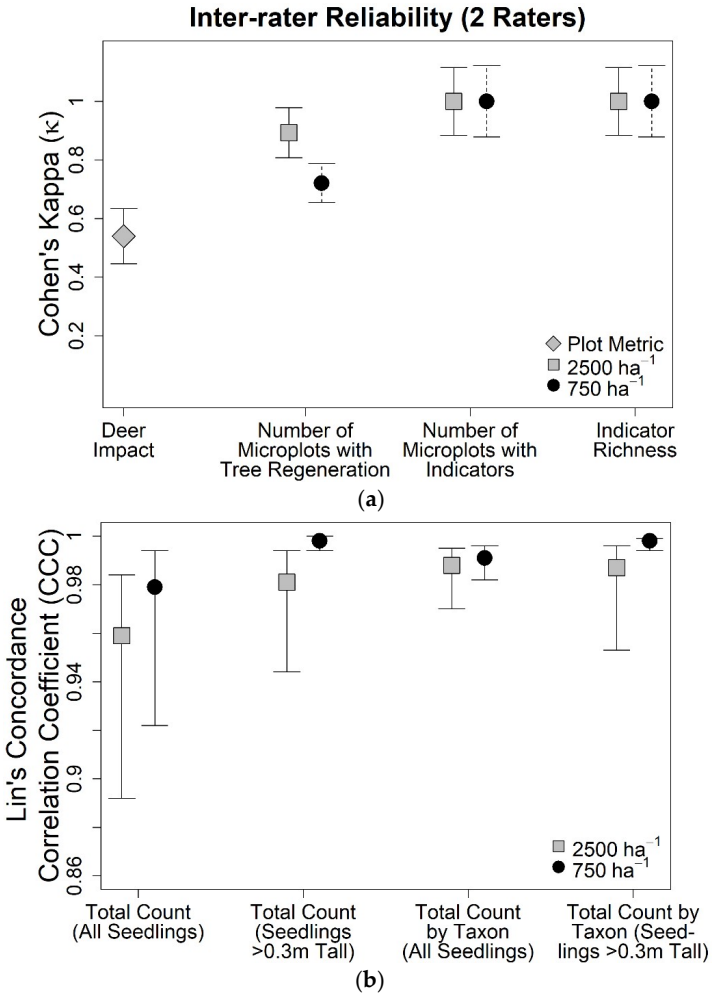


Figure 4. Inter-rater reliability (IRR) assessments of 4 categorical vegetation metrics (**a**) compared with Cohen's Kappa (κ), and 4 continuous vegetation metrics (**b**) compared with Lin's concordance correlation coefficient (CCC). Gray squares indicate values for the 1/2500th ha scale, black circles indicate values for the 1/750th ha scale, and the gray diamond indicates a plot-level metric. Dashed and gray lines represent 95% confidence intervals for 1/750th and 1/2500th ha point values, respectively.

Power was high (≥0.90) only when the sample size was ≥40 for the largest mean stem change (2.0 stems per plot), and differences in stem counts between the time 1 and time 2 data sets were detected, on average, at a rate of 95% for the 1/2500th ha scale and a rate of 91% for the 1/750th ha scale (Figure 5). For a 1.5 mean stem change, power was comparable and equaled or exceeded 90% when the sample size was ≥50, regardless of scale. For a 1.0 stem change, power reached 90% when sample size was 60 for the 1/2500th ha scale and 80 for the 1/750th ha scale. Overall, power was moderate (<0.84) for the 0.5 mean stem change for both scales of data collection, and differences in stem counts were only detected at a rate of 84% for the 1/2500th ha scale and 76% for the 1/750th ha scale for the largest sample size (100). Generally, there was increased power to detect change in mean stem counts at the 1/2500th ha scale compared to the 1/750th ha scale, regardless of the amount of change modeled.

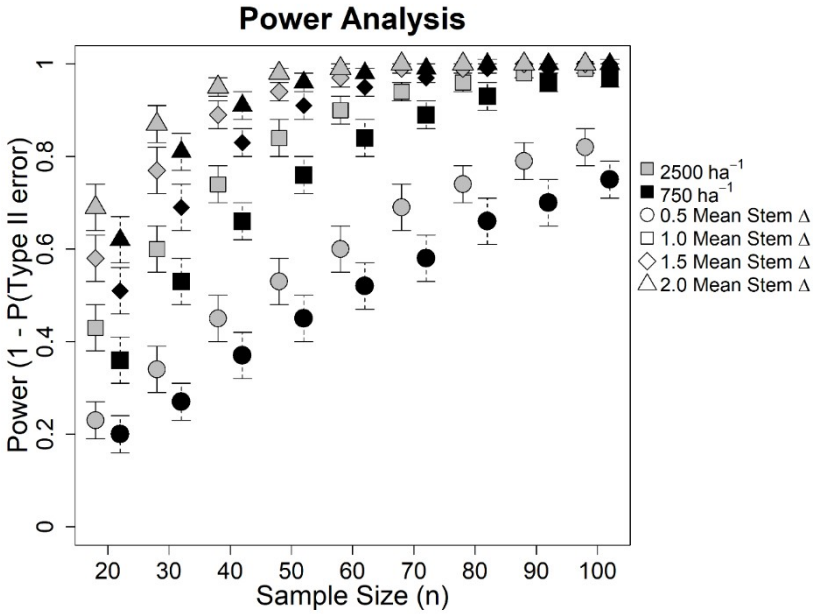


Figure 5. Statistical power to detect mean tree seedling stem change across sample size (20 to 100) and scale (1/2500th ha and 1/750th ha) based on simulated tree seedling stem counts with added sampling (observer error). Shapes indicate different levels of mean tree seedling stem change added to initial plot stem counts (circle = 0.5, square = 1.0, diamond = 1.5, and triangle = 2.0), while color indicates scale (gray = 1/2500th ha, and black = 1/750th ha). Dashed black lines and solid gray lines represent the standard deviation of each point estimate for the 1/2500th and 1/750th ha, respectively.

4. Discussion

Vegetation monitoring programs with pre-survey training contribute to increased data consistency [18]. Following training, our field crews were consistent in their evaluation of quantifiable, measurable habitat characteristics including counts of tree seedlings, counts of microplots with tree regeneration or indicator plants, and indicator richness. The same was true with counts of microplots with tree regeneration at the 1/2500th ha scale, but there was a significant reduction in reliability at the 1/750th ha scale demonstrating higher rates of observer error even with extensive training on data collection protocols. Of all the categorical metrics evaluated, the DII was the least reliable (there was only a moderate level of agreement between field crews in their DII scores), suggesting that data derived from actual field measurements is inherently more dependable than subjective indices derived from general plot observations. The difference in reliability across measured and subjective categories is seen in other disciplines, in which subjective categorical indices have lower measures of inter-rater reliability [52–56].

Counts of tree seedlings, regardless of count type, were reliable between field crews. Agreement was substantial to almost perfect, and there was no significant difference between reliability across scale, supporting the use of tree seedling counts as a consistent way to track changes in tree regeneration over time. Furthermore, crews consistently counted tree seedlings regardless of height or taxon, indicating that both tree seedling identification and height class counts were comparable. Reliability of tree seedling counts by taxon suggests that changes in species composition and/or seedling height classes could also be monitored in addition to overall seedling counts to consistently evaluate shifts in tree regeneration as a part of a long-term monitoring program. Plant communities are complex entities, and taxa within a community rarely respond uniformly to changes in their environment [57,58]. The availability of several metrics to evaluate changes in tree seedling regeneration could better capture the mechanisms behind these shifts or provide a more complete picture of tree seedling responses [30,59,60]. In sum, there was substantial inter-rater reliability across all vegetation metrics yielding confidence in the consistency of data collection across field crews. Additionally, there was no support for increased reliability of vegetation survey metrics at smaller sampling scales (H1), because there was no difference in inter-rater reliability across scale for 6 of the 7 metrics evaluated.

As expected, power to detect changes in tree seedling stem counts increased with increasing sample size and magnitude of change modeled [61]. Across each mean stem change, statistical power was greater at the smaller sampling scale than the larger one, signaling a positive relationship between observer error and scale. Reductions in statistical power across scale are exclusively attributed to observer error because mathematically, only the error term varied across scale in the power analysis. A reduction in statistical power signals that there was more variability in stem counts between crews at the larger scale, supporting our hypothesis (H2). Our study design does not limit survey time at each location, because limitations on survey time have been shown to increase observer error [62]. Instead, the increased probability of finding more species, combined with higher cumulative counts of abundant vegetation (including seedlings), likely explains increases in observer error at the 1/750th ha sampling scale. Crews are more likely to miss or double count stems in dense vegetation, especially when seedling abundance is high. Overall, the effectiveness of the smaller 1/2500th ha plot size at detecting changes in tree seedling counts indicates that it is an adequate scale of data collection for our study. Despite studies quantifying observer error as a part of long-term vegetation monitoring or citizen science projects [3,18,63,64], the incorporation of observer error into the analysis of plant data is a relatively new technique in plant ecology [65,66]. To our knowledge, despite past studies assessing the effects of plot size on ordination techniques [67], variances in basal area estimates [68], or species constancy and species richness [69,70], this is the first study to evaluate the effect of scale on power to detect changes in understory vegetation.

The power analysis suggests that a relatively large sample size ($n \geq 40$) is needed to detect a modest 2.0 stem change in tree seedling counts per plot. This threshold is above the general ecological rule of thumb of 30 sampling sites for ecological studies [36,71], suggesting high inter-plot variability in stem counts across the study area. This is not surprising given the wide geographic area covered (~100 km^2), and the variation in habitat types encompassed by our random design [72]. However, past CFI data suggests that most changes in mean stem counts between sampling intervals across all state forest lands are less than 2.0 stems per plot (Figure A1), suggesting that even larger sample sizes ($n \geq 60$) may be required to detect differences in tree seedling counts over time. Due to the inherent variability in vegetation characteristics across sampling sites, the effect of observer error on detecting changes in understory vegetation is amplified. Reducing observer error and sampling at smaller scales will improve inter-rater reliability and statistical power.

5. Conclusions

Vegetation monitoring programs designed to intensely monitor understory vegetation data (including seedlings and herbaceous plants) require consistent data collection methods to monitor changes to understory plant diversity and abundance [62]. As this study demonstrates, smaller sampling scales can improve statistical power through improved consistency in data collection, but the required sample size to detect change was high ($n > 40$). The use of 1/2500th ha plots and objective, reliable vegetation metrics (like tree seedling counts and the presence of phyto-indicators) suggests that the Bureau of Forestry's VIP is likely to detect changes in deer impact in areas enrolled in DMAP. However, the PGC's reliance on the FIA DII to make deer management decisions across larger geographic areas may prove problematic, because the DII is a subjective measure prone to only moderate levels of inter-rater reliability. Revising the DII to incorporate more reliable vegetation metrics may improve its inter-rater reliability, but it is more likely that objective vegetation measures (like FIA tree seedling stem counts) will better detect changes in deer impact across PGC wildlife management units. Revising decision models to incorporate direct FIA vegetation measures would help the PGC make better deer management decisions that align with their objectives.

The relatively high levels of inter-rater reliability between field crews across both categorical and continuous (count) data metrics highlight the importance of quantifiable, measurement-based data collection protocols. Alternatively, the reliability of the only subjective, observational metric (the DII) was low, raising concerns about its ability to track changes in deer browsing effects through time. Our large, landscape-level sampling areas required 10 additional sampling locations than recommended for ecological studies due to the high variability in vegetation conditions between sampling locations [71]. This heterogeneity among sampling locations indicates that other studies of similar size and vegetation composition may require larger sample sizes to detect vegetation changes through time. Identifying specific project objectives and designing sampling protocols to achieve those objectives is recommended for other studies considering long-term ecological monitoring of vegetation conditions to track responses to changes in deer browsing pressure.

Author Contributions: Conceptualization, D.R.B.-M., D.R.D., M.E.M., C.S.R. and E.H.J.; Methodology, D.R.B.-M., D.R.D. and M.E.M.; Validation, D.R.D. and M.E.M.; Formal Analysis, D.R.B.-M. and D.R.D.; Investigation, D.R.B.-M.; Resources, D.R.D., M.E.M., C.S.R. and E.H.J.; Data Curation, D.R.B.-M.; Writing—Original Draft Preparation, D.R.B.-M.; Writing—Review & Editing, D.R.B.-M., D.R.D., M.E.M., C.S.R. and E.H.J.; Visualization, D.R.B.-M. and D.R.D.; Supervision, D.R.D. and M.E.M.; Project Administration, D.R.B.-M.; Funding Acquisition, D.R.D., M.E.M., C.S.R. and E.H.J.

Funding: This was a Pennsylvania Cooperative Fish and Wildlife Research Unit research project funded by the Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry, the Pennsylvania Game Commission, and the Pennsylvania State University.

Acknowledgments: Special thanks to K. Duren with the DCNR Bureau of Forestry for providing sampling means and variances from VIP sampling data, T. Albright with the USDA Forest Service for help training field crews, and the field technicians who worked to collect vegetation inventory data: N. Wilson, M. McArthur, M. Rothrock, S. Gaffney, and M. Antonishak. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A

Table A1. List of tree seedling taxa considered desirable regeneration by the Department of Conservation and Natural Resources Bureau of Forestry, of which nearly all species were browse-preferred by white-tailed deer.

Common Name	Scientific Name
American basswood	*Tilia americana* L.
Bigtooth aspen	*Populus grandidentata* Michx.
Black ash	*Fraxinus nigra* Marshall
Black cherry	*Prunus serotina* Ehrh.
Black oak	*Quercus velutina* Lam.
Chestnut oak	*Quercus montana* Willd.
Cucumbertree	*Magnolia acuminata* L.
Green ash	*Fraxinus pennsylvanica* Marshall
Hemlock	*Tsuga canadensis* (L.) Carrière
Hickory (genus)	*Carya* spp.
Paper birch	*Betula papyrifera* Marshall
Pitch pine	*Pinus rigida* Mill.
Quaking aspen	*Populus tremuloides* Michx.
Red maple	*Acer rubrum* L.
Red oak	*Quercus rubra* L.
Scarlet oak	*Quercus coccinea* Münchh.
Sugar maple	*Acer saccharum* Marshall
White ash	*Fraxinus americana* L.
White oak	*Quercus alba* L.
White pine	*Pinus strobus* L.
Yellow birch	*Betula alleghaniensis* Britton
Yellow poplar	*Liriodendron tulipifera* L.

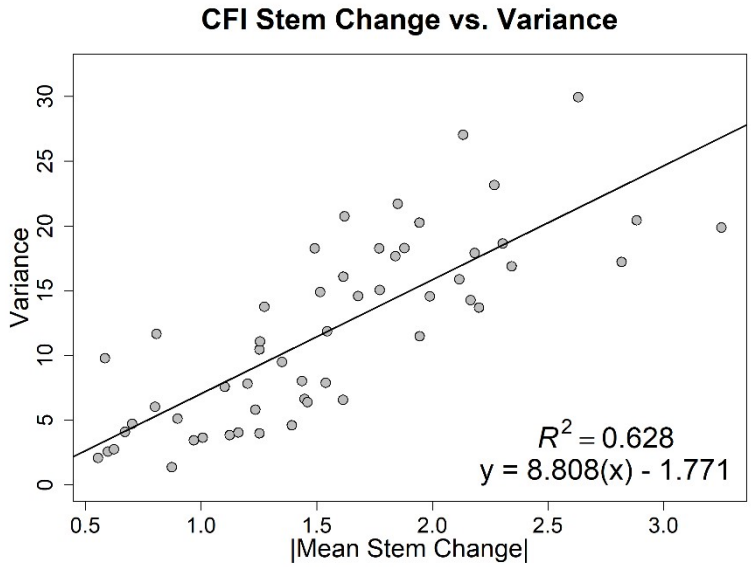


Figure A1. Plot of Vegetation Impact Protocol (VIP) absolute value mean stem change and sampling variance for 19 state forest districts across 3 sampling periods. The best-fit linear trendline is plotted and the regression equation is also listed.

References and Notes

1. Lindenmayer, D.B.; Likens, G.E.; Andersen, A.; Bowman, D.; Bull, C.M.; Burns, E.; Dickman, C.R.; Hoffmann, A.A.; Keith, D.A.; Liddell, M.J.; et al. Value of long-term ecological studies. *Austral Ecol.* **2012**, *37*, 745–757. [CrossRef]

2. Ravlin, F.W.; Voshell, J.R., Jr.; Smith, D.W.; Rutherford, S.L.; Hiner, S.W.; Haskell, D.A. Section I: Overview. In *Shenandoah National Park Long-Term Ecological Monitoring System User Manuals*; U.S. Department of the Interior, National Park Service: Washington, DC, USA, 1990; pp. I-1–I-17.
3. Milberg, P.; Bergstedt, J.; Fridman, J.; Odell, G.; Westerberg, L. Observer bias and random variation in vegetation monitoring data. *J. Veg. Sci.* **2008**, *19*, 633–644. [CrossRef]
4. Symstad, A.J.; Wienk, C.L.; Thorstenson, A.D. Precision, repeatability, and efficiency of two canopy-cover estimate methods in northern great plains vegetation. *Rangel. Ecol. Manag.* **2008**, *61*, 419–429. [CrossRef]
5. Chandra Sekar, K.; Rawal, R.S.; Chaudhery, A.; Pandey, A.; Rawat, G.; Bajapai, O.; Joshi, B.; Bisht, K.; Mishra, B.M. First GLORIA site in Indian Himalayan region: Towards addressing issue of long-term data deficiency in the Himalaya. *Natl. Acad. Sci. Lett.* **2017**, *40*, 355–357. [CrossRef]
6. Fleming, G.M.; Diffendorfer, J.E.; Zedler, P.H. The relative importance of disturbance and exotic-plant abundance in California coastal sage scrub. *Ecol. Appl.* **2009**, *19*, 2210–2227. [CrossRef] [PubMed]
7. Li, S.; Pennings, S.C. Disturbance in Georgia salt marshes: Variation across space and time. *Ecosphere* **2016**, *7*, 1–11. [CrossRef]
8. Van Lierop, P.; Lindquist, E.; Sathyapala, S.; Franceschini, G. Global forest area disturbance from fire, insect pests, diseases and severe weather events. *For. Ecol. Manag.* **2015**, *352*, 78–88. [CrossRef]
9. Melillo, J.M.; Butler, S.; Johnson, J.; Mohan, J.; Steudler, P.; Lux, H.; Burrows, E.; Bowles, F.; Smith, R.; Scott, L.; et al. Soil warming, carbon-nitrogen interactions, and forest carbon budgets. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9508–9512. [CrossRef] [PubMed]
10. Mahan, C.G.; Diefenbach, D.R.; Cass, W.B. Evaluating and revising a long-term monitoring program for vascular plants: Lessons from Shenandoah National Park. *Nat. Areas J.* **2007**, *27*, 16–24. [CrossRef]
11. Munson, S.M.; Duniway, M.C.; Johanson, J.K. Rangeland monitoring reveals long-term plant responses to precipitation and grazing at the landscape scale. *Rangel. Ecol. Manag.* **2016**, *69*, 76–83. [CrossRef]
12. Bagchi, S.; Singh, N.J.; Briske, D.D.; Bestelmeyer, B.T.; McClaran, M.P.; Murthy, K. Quantifying long-term plant community dynamics with movement models: Implications for ecological resilience: Implications. *Ecol. Appl.* **2017**, *27*, 1514–1528. [CrossRef] [PubMed]
13. Frerker, K.L.; Sabo, A.; Waller, D. Long-term regional shifts in plant community composition are largely explained by local deer impact experiments. *PLoS ONE* **2014**, *9*, e0185037. [CrossRef] [PubMed]
14. Groffman, P.M.; Rustad, L.E.; Templer, P.H.; Campbell, J.L.; Lynn, M.; Lany, N.K.; Socci, A.M.; Vadeboncoeur, M.A.; Schaberg, P.G.; Wilson, F.; et al. Long-term integrated studies show complex and surprising effects of climate change in the northern hardwood forest. *Bioscience* **2012**, *62*, 1056–1066. [CrossRef]
15. Liknes, G.C.; Nelson, M.D.; Kaisershot, D.J. *Net Change in Forest Density, 1873–2001: Using Historical Maps to Monitor Long-Term Forest Trends*; U.S. Department of Agriculture, Forest Service, Northern Research State: Newton Square, PA, USA, 2011; p. 12.
16. Miles, P.D.; Brand, G.J.; Alerich, C.L.; Bednar, L.F.; Woudenberg, S.W.; Glover, J.F.; Ezzell, E.N. *The Forest Inventory and Analysis Database: Database Description and Users Manual*; version 1.0; U.S. Department of Agriculture, Forest Service, Northern Central Reseach Station: St. Paul, MN, USA, 2001.
17. Goeking, S.A. Disentangling Forest change from forest inventory change: A case study from the US Interior West. *J. For.* **2015**, *113*, 475–483. [CrossRef]
18. Morrison, L.W. Observer error in vegetation surveys: A review. *J. Plant Ecol.* **2016**, *9*, 367–379. [CrossRef]
19. Bosworth, D. *Forest Inventory and Analysis Strategic Plan*; U.S. Department of Agriculture, Forest Service: Washington, DC, USA, 2007.
20. U.S. Department of Agriculture, Forest Service, Northern Research Station. Forest Inventory and Analysis National Core Field Guide—Volume I Supplement: Field Data Collection Procedures For Phase 2+ Plots. Available online: https://www.nrs.fs.fed.us/fia/data-collection/field-guides/ver7.1/NRS%20FG%207.1-April%202017-Complete%20Document_NRSP2plus.pdf (accessed on 15 October 2018).
21. Department of Conservation and Natural Resources Bureau of Forestry. Inventory Manual of Procedure for The 4th Cycle of CFI Measurements (2015–2020), Inventory of Biological Resources.
22. U.S. Department of Agriculture, Forest Service. About the Agency: What We Believe. Available online: https://www.fs.fed.us/about-agency (accessed on 18 October 2018).

23. Pennsylvania Department of Conservation and Natural Resources. DCNR Bureau of Forestry—Our Mission and What We Do. Available online: https://www.dcnr.pa.gov/about/Pages/Forestry.aspx (accessed on 18 October 2018).
24. Albright, T.A.; McWilliams, W.H.; Widmann, R.H.; Butler, B.J.; Crocker, S.J.; Kurtz, C.M.; Lehman, S.L.; Lister, T.W.; Miles, P.D.; Morin, R.S.; et al. *Pennsylvania Forests 2014*; U.S. Department of Agriculture, Forest Service, Northern Research Station: Newton Square, PA, USA, 2017.
25. Pennsylvania Department of Conservation and Natural Resources, Bureau of Forestry. Detecting & Monitoring Vegetation Changes within DMAP Units: Vegetation Impact Protocol in an Adaptive Resource Management Context.
26. Pennsylvania—Department of Conservation and Natural Resource—Bureau of Forestry—Ecological Service. White-Tailed Deer Plan. Available online: http://www.docs.dcnr.pa.gov/cs/groups/public/documents/document/dcnr_20027101.pdf (accessed on 15 October 2018).
27. Begley-Miller, D.R.; Hipp, A.L.; Brown, B.H.; Hahn, M.; Rooney, T.P. White-tailed deer are a biotic filter during community assembly, reducing species and phylogenetic diversity. *AoB Plants* **2014**, *6*, 1–9. [CrossRef] [PubMed]
28. Rooney, T.P. High white-tailed deer densities benefit graminoids and contribute to biotic homogenization of forest ground-layer vegetation. *Plant Ecol.* **2009**, *202*, 103–111. [CrossRef]
29. Côté, S.D.; Rooney, T.P.; Tremblay, J.-P.; Dussault, C.; Waller, D.M. Ecological impacts of deer overabundance. *Annu. Rev. Ecol. Evol. Syst.* **2004**, *35*, 113–147. [CrossRef]
30. Habeck, C.W.; Schultz, A.K. Community-level impacts of white-tailed deer on understorey plants in North American forests: A meta-analysis. *AoB Plants* **2015**, *7*, plv119. [CrossRef] [PubMed]
31. Tilghman, N.G. Impacts of white-tailed deer on forest regeneration in northwestern Pennsylvania. *J. Wildl. Manag.* **1989**, *53*, 524–532. [CrossRef]
32. Horsley, S.B.; Stout, S.L.; DeCalesta, D.S. White-tailed deer impact on the vegetation dynamics of a northern hardwood forest. *Ecol. Appl.* **2003**, *13*, 98–118. [CrossRef]
33. Rosenberry, C.S.; Fleegle, J.T.; Wallingford, B.D. *Management and Biology of White-Tailed Deer in Pennsylvania 2009–2018*; Pennsylvania Game Commission: Harrisburg, PA, USA, 2009.
34. McCarthy, M.A.; Moore, J.L.; Morris, W.K.; Parris, K.M.; Garrard, G.E.; Vesk, P.A.; Rumpff, L.; Giljohann, K.M.; Camac, J.S.; Bau, S.S.; et al. The influence of abundance on detectability. *Oikos* **2013**, *122*, 717–726. [CrossRef]
35. Müller, F.; Baessler, C.; Schubert, H.; Klotz, S. *Long-Term Ecological Research: Between Theory and Application*; Springer: Dordecht, The Netherlands, 2010; ISBN 9789048187812.
36. Martínez-Abraín, A. Is the "n = 30 rule of thumb" of ecological field studies reliable? A call for greater attention to the variability in our data. *Anim. Biodivers. Conserv.* **2014**, *37*, 95–100.
37. Russell, M.B.; Woodall, C.W.; Potter, K.M.; Walters, B.F.; Domke, G.M.; Oswalt, C.M. Interactions between white-tailed deer density and the composition of forest understories in the northern United States. *For. Ecol. Manag.* **2017**, *384*, 26–33. [CrossRef]
38. Beardall, V.; Gill, R.M.A. The impact of deer on woodlands: The effects of browsing and seed dispersal on vegetation structure and composition. *Forestry* **2001**, *74*, 209–218.
39. Frerker, K.L.; Sonnier, G.; Waller, D.M. Browsing rates and ratios provide reliable indices of ungulate impacts on forest plant communities. *For. Ecol. Manag.* **2013**, *291*, 55–64. [CrossRef]
40. Waller, D.M.; Alverson, W.S. The White-tailed deer: A keystone herbivore. *Wildl. Soc. Bull.* **1997**, *25*, 217–226.
41. National Oceanic and Atmospheric Administration. Normal Dates of Last Freeze in Spring and First Freeze in Autumn Across Central Pennsylvania. Available online: http://www.weather.gov/ctp/FrostFreeze (accessed on 1 January 2017).
42. National Oceanic and Atmospheric Administration NOWData NOAA Data Online Weather Data. Available online: http://w2.weather.gov/climate/xmacis.php?wfo=ctp (accessed on 1 January 2017).
43. Kirschbaum, C.D.; Anacker, B.L. The utility of Trillium and Maianthemum as phyto-indicators of deer impact in northwestern Pennsylvania, USA. *For. Ecol. Manag.* **2005**, *217*, 54–66. [CrossRef]
44. Rooney, T.P.; Gross, K. A demographic study of deer browsing impacts on *Trillium grandiflorum*. *Plant Ecol.* **2003**, *168*, 267–277. [CrossRef]
45. Royo, A.A.; Stout, S.L.; DeCalesta, D.S.; Pierson, T.G. Restoring forest herb communities through landscape-level deer herd reductions: Is recovery limited by legacy effects? *Biol. Conserv.* **2010**, *143*, 2425–2434. [CrossRef]

46. Gamer, M.; Lemon, J.; Fellows, I.; Singh, P. Package 'irr': Various Coefficients of Interrater Reliability and Agreement. Available online: https://cran.r-project.org/web/packages/irr/irr.pdf (accessed on 15 October 2018).
47. R Core Team. R: A Language and Environment for Statistical Computing. Available online: https://www.R-project.org (accessed on 15 October 2018).
48. Mabmud, S.M. Cohen's Kappa. In *Encyclopedia of Research Design*; Salkind, N., Ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2012; pp. 188–189, ISBN 9781412961271.
49. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Int. Biom. Soc.* **1977**, *33*, 159–174. [CrossRef]
50. Carrasco, J.L.; Martinez, J.P. Package 'cccrm': Concordance Correlation Coefficient for Repeated (and Non-Repeated) Measures. Available online: https://cran.r-project.org/web/packages/cccrm/cccrm.pdf (accessed on 15 October 2018).
51. King, T.S.; Cinchilli, V.M.; Carrasco, J.L. A repeated measures concordance correlation coefficient. *Stat. Med.* **2007**, *26*, 3095–3113. [CrossRef] [PubMed]
52. Lin, L.I.; McBride, G.; Bland, J.M.; Altman, D.G. *A Proposal For Strength-Of-Agreement Criteria for Lin'S Concordance Correlation Coefficient*; National Institute of Water & Atmopheric Research Ltd.: Hamilton, New Zealand, 2005.
53. Zanarini, M.C.; Frankenburg, F.R.; Vujanovic, A. Inter-rater and test-retest reliability of the Revised Diagnostic Interview for Borderlines. *J. Pers. Disord.* **2002**, *16*, 270–276. [CrossRef] [PubMed]
54. Awatani, T.; Morikita, I.; Shinohara, J.; Mori, S.; Nariai, M.; Tatsumi, Y.; Nagata, A.; Koshiba, H. Intra- and inter-rater reliability of isometric shoulder extensor and internal rotator strength measurements performed using a hand-held dynamometer. *J. Phys. Ther. Sci.* **2016**, *28*, 3054–3059. [CrossRef] [PubMed]
55. Meeremans, P.; Yochum, N.; Kochzius, M.; Ampe, B.; Tuyttens, F.A.M.; Uhlmann, S.S. Inter-rater reliability of categorical versus continuous scoring of fish vitality: Does it affect the utility of the reflex action mortality predictor (RAMP) approach? *PLoS ONE* **2017**, *12*, e0179092. [CrossRef] [PubMed]
56. Björk, J.; Rittner, R.; Cromley, E. Exploring inter-rater reliability and measurement properties of environmental ratings using kappa and colocation quotients. *Environ. Health* **2014**, *13*, 1–11. [CrossRef] [PubMed]
57. Smith, S.D.; Charlet, T.N.; Zitzer, S.F.; Abella, S.R.; Vanier, C.H.; Huxman, T.E. Long-term response of a Mojave Desert winter annual plant community to a whole-ecosystem atmospheric CO_2 manipulation (FACE). *Glob. Chang. Biol.* **2014**, *20*, 879–892. [CrossRef] [PubMed]
58. Ritchie, M.E.; Tilman, D.; Knops, J.M.H. Herbivore effects on plant and nitrogen dynamics in oak savanna. *Ecology* **1998**, *79*, 165–177. [CrossRef]
59. Morrissey, R.C.; Jacobs, D.F.; Seifert, J.R. *Response of Northern Red Oak, Black Walnut, and White Ash Seedlings to Various Levels of Simulated Summer Deer Browsing*; U.S. Department of Agriculture, Forest Service, Northern Research Station: Newtown Square, PA, USA, 2008; pp. 50–58.
60. Augustine, D.J.; McNaughton, S.J. Ungulate effects on the functional species composition of plant communities: Herbivore selectivity and plant tolerance. *J. Wildl. Manag.* **1998**, *62*, 1165–1183. [CrossRef]
61. Akobeng, A.K. Understanding type I and type II errors, statistical power and sample size. *Acta Paediatr.* **2016**, *105*, 605–609. [CrossRef] [PubMed]
62. Zhang, J.; Nielsen, S.E.; Grainger, T.N.; Kohler, M.; Chipchar, T.; Farr, D.R. Sampling plant diversity and rarity at landscape scales: Importance of sampling time in species detectability. *PLoS ONE* **2014**, *9*. [CrossRef] [PubMed]
63. Butt, N.; Slade, E.; Thompson, J.; Malhi, Y.; Riutta, T. Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. *Ecol. Appl.* **2013**, *23*, 936–943. [CrossRef] [PubMed]
64. Vittoz, P.; Guisan, A. How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *J. Veg. Sci.* **2017**, *18*, 413–422. [CrossRef]
65. Wright, W.J.; Irvine, K.M.; Warren, J.M.; Barnett, J.K. Statistical design and analysis for plant cover studies with multiple sources of observation errors. *Methods Ecol. Evol.* **2017**, *8*, 1832–1841. [CrossRef]
66. Mason, N.W.H.; Holdaway, R.J.; Richardson, S.J. Incorporating measurement error in testing for changes in biodiversity. *Methods Ecol. Evol.* **2018**, *9*, 1296–1307. [CrossRef]

67. Otypková, Z.; Chytrý, M. Effects of plot size on the ordination of vegetation samples. *J. Veg. Sci.* **2006**, *17*, 465–472. [CrossRef]
68. Bormann, F.H. The statistical efficience of the sample plot. *Ecology* **1953**, *34*, 474–487. [CrossRef]
69. Dengler, J.; Löbel, S.; Dolnik, C. Species constancy depends on plot size—A problem for vegetation classification and how it can be solved. *J. Veg. Sci.* **2009**, *20*, 754–766. [CrossRef]
70. Stohlgren, T.J.; Chong, G.W.; Kalkhan, M.A.; Schell, L.D.; Applications, S.E.; Aug, N. Multiscale Sampling of Plant Diversity: Effects of minimum mapping unit size. *Ecol. Appl.* **2014**, *7*, 1064–1074. [CrossRef]
71. Johnson, S.E.; Mudrak, E.L.; Beever, E.A.; Sanders, S.; Waller, D.M. Comparing power among three sampling methods for monitoring forest vegetation. *Can. J. For. Res.* **2008**, *38*, 143–156. [CrossRef]
72. Fike, J. (Ed.) *Terrestrial & Palustrine Plant Communities of Pennsylvania*; Pennsylvania Department of Conservation and Natural Resources: Harrisburg, PA, USA, 1999.

This article is published under the terms of the free Open Government License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited: See: https://www.copyrightlaws.com/copyright-laws-in-u-s-government-works/.