

Power Analysis of the Ecological Monitoring Network

Amy E. Kendig, Nathan Dahlberg, and Holly Bernardo

March 7, 2023

Cover photo credit: Katie Reisinger

Division of Ecological and Water Resources
500 Lafayette Road
St. Paul, MN 55155-4040
(651) 296-6157 or (888) 646-6367

The Minnesota DNR prohibits discrimination in its programs and services based on race, color, creed, religion, national origin, sex, marital or familial status, disability, public assistance status, age, sexual orientation, and local human rights commission activity. Individuals with a disability who need a reasonable accommodation to access or participate in DNR programs and services, including those who would like to request this document in an alternative format, should contact the DNR ADA Title II Coordinator at info.dnr@state.mn.us or 651-296-6157. We welcome calls from Telecommunications Relay Service (TRS) users. For assistance in other languages, please call 651-296-6157 or 888-MINNDNR (646-6367). Discrimination inquiries should be sent to Minnesota DNR, 500 Lafayette Road, St. Paul, MN 55155-4049; or U.S. Environmental Protection Agency, Mail code 1201A, 1200 Pennsylvania Avenue, MW, Washington, DC 2046.

Thank you to David Minor for help with preparing the final report and to the Ecological Monitoring Network field surveyors for the data used in this report. Funding for the Ecological Monitoring Network was provided by the Minnesota Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR).

The document should be cited as:

Kendig, A. E., N. Dahlberg, and H. Bernardo. 2023. Power Analysis of the Ecological Monitoring Network. Minnesota Department of Natural Resources, St. Paul, MN. 48 pp.

For comments and questions about this document, please contact:

Amy Kendig
Biometrician/Wetland Research Scientist
Minnesota Biological Survey
amy.kendig@state.mn.us
(651) 259-5116

© 2026, State of Minnesota, Department of Natural Resources

Contents

Power Analysis of the Ecological Monitoring Network 1

Contents 3

Executive summary 4

Introduction 5

Methods 6

Overview 6

Inputs 7

Simulating data 10

Fitting models 12

Analyses with random and non-random plots 13

Results 14

Statewide analyses 14

Community analyses 25

Forest impacts 32

Appendices 38

Appendix A: Sample Selection 38

Appendix B: Analyzing Community Composition 39

Appendix C: Generic indicators with alternative baseline averages 40

References 44

Executive summary

We evaluated study designs of the Ecological Monitoring Network (EMN), assessing their potential to estimate trends and impacts with sufficient accuracy and precision. Study designs varied in plot number and revisit frequency. We evaluated six indicators of statewide biodiversity trends: native and non-native ground layer plant richness and cover, graminoid cover, and a generic undefined univariate indicator. We evaluated four indicators of community-scale biodiversity trends: shifts in community composition, biotic homogenization within a community, biotic homogenization between communities, and a generic undefined univariate indicator. Finally, we evaluate three indicators of forest impacts: the effect of deer browsing on *Acer* spp. occurrence, the effect of earthworm disturbance on graminoid cover, and the effect of timber harvesting on *Acer* spp. basal area. For four of the indicators–statewide native ground layer plant richness, statewide graminoid cover, the community-scale generic indicator, and the effect of deer browsing–we evaluated the effect of including non-randomly selected plots with slightly different ecology from randomly selected plots. Together, these analyses provided recommendations on the number of plots and the proportion of plots to be chosen non-randomly in upcoming field seasons for EMN.

To detect trends in statewide biodiversity indicators with estimates within 0.5% of the true value and with >87% power, a conservative recommendation is 400 plots and four revisits. This sample size was primarily derived from the generic indicator, which had relatively high spatial variation (equal to that for non-native ground layer plant cover). However, this recommended sample size would only meet the stated criteria for strong trends (i.e., 10% annual increase or decrease). Subtler trends (+/-5%) could be detected with 400 plots with high accuracy and precision when spatial variation was lower, such as for native ground layer plant richness or cover, or non-native ground layer plant richness. The large sample sizes for statewide analyses generally protected estimates from inaccuracy and imprecision introduced by purposely selecting plots that had different trends than the randomly selected plots. If 600 total plots are completed, maintaining three-quarters of the plots as random is recommended for ensuring greater accuracy (coverage >77%, based on graminoid cover analysis).

To detect trends in community-scale biodiversity indicators with estimates within 0.5% of the true value and with >87% power, fifty plots and four revisits are recommended. This sample size was based on the generic indicator with the statewide spatial variation divided by ten. Fewer plots were required for indicators with more consistent measurements and trends across plots. To detect shifts in community composition with >87% power, 20 plots and four revisits were recommended. Biotic homogenization can be estimated within 0.5% of the true value and with >87% power with 20 plots and one revisit. It is recommended that all plots in these community-scale analyses are selected randomly to maintain acceptable precision.

Analyses of forest impacts suggested that 200 plots are needed to obtain effect estimates within 0.5% of the true value and that have >87% power. This sample size applied to effect sizes of +/-5%, which was consistent with effect sizes measured for deer browse and earthworm disturbance in the existing dataset. However, earthworm disturbance effects will be generally difficult to detect (power never exceeded 60% with 200 plots), likely due to the ordinal nature of the disturbance variable. Similarly, the recommended sample size of 200 plots only met the accuracy and precision criteria for timber harvest effect sizes of +/-10%. Careful consideration of how to quantify drivers of impact may increase our ability to detect effects.

Introduction

The Ecological Monitoring Network (EMN) is a statewide monitoring program that tracks long-term vegetation changes in forests, prairies, and wetlands across Minnesota. Potential drivers of change include climate change, invasive species, disease, loss of pollinators, habitat fragmentation, land-use change, and use of water resources. The network of EMN plots is intended to represent Minnesota's native plant communities. Researchers have collected data from hundreds of plots throughout the state and are now at a junction to decide how many more plots are needed to address the program's goals. The objective of this report is to evaluate how many plots (i.e., sample size) are needed to achieve the program's goals with sufficient accuracy and precision.

We sought to evaluate sample sizes needed to estimate:
1. Trends in statewide biodiversity indicators
2. Trends in plant community biodiversity indicators
3. The effects of deer, earthworms, and timber harvesting in forested plant communities

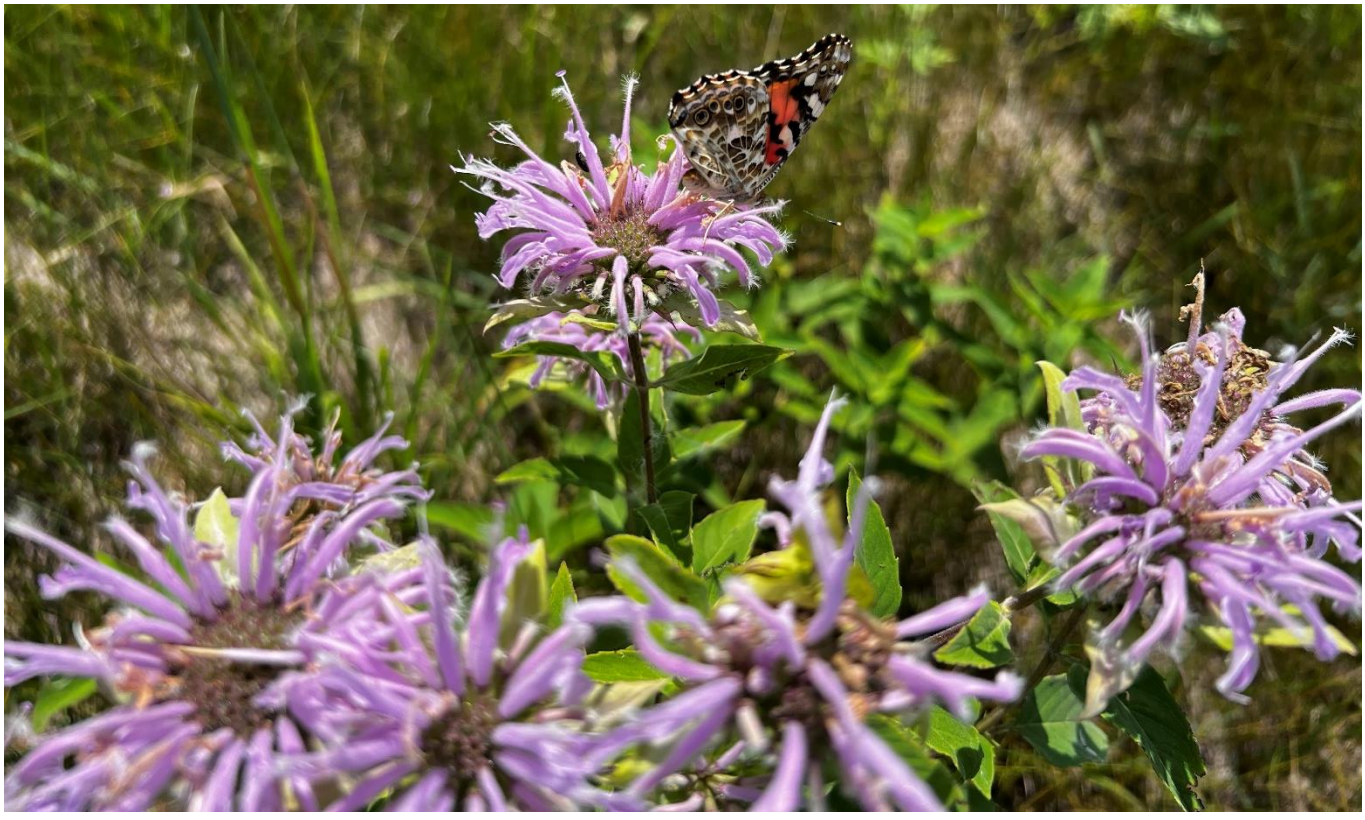


Photo credit: Katie Reisinger

Methods

Overview

Statistical power is the probability that a test will correctly reject the null hypothesis (e.g., no change over time in indicator) when a specific alternative hypothesis is true (e.g., 1% annual increase in indicator). Power depends on the sample size, magnitude of change of the indicator, variation in the indicator and its change over time, plot revisit schedule, duration of monitoring program, statistical model, and many other factors (Urquhart and Kincaid 1999, Johnson et al. 2008, Irvine and Rodhouse 2010, Miller and Mitchell 2014, Perles et al. 2014b, Starcevich et al. 2018).

One approach to evaluating power is to simulate datasets with characteristics of interest. These may be closely based on data already collected or drawn from distributions with parameters from the literature. The datasets are created to adhere to a specific alternative hypothesis. The statistical test is applied to each dataset and the proportion of datasets for which the null hypothesis is correctly rejected is the power (Bolker 2008).

For each indicator, study design, and trend value, we generated 500 simulated datasets and fit a statistical model to each one. The simulated datasets have properties that we expect will be similar to data collected from EMN. In addition to evaluating power, we also assessed bias, confidence interval (CI) width, and coverage (Bolker 2008) (Table 1).

Table 1: Metrics used to evaluate sample sizes.

Metric	Measure	Description
bias	accuracy	difference between the estimate and true value (i.e., the trend or effect size used to generate the simulate data), averaged across simulated datasets
CI width	precision	difference between the estimated upper and lower 95% confidence intervals, averaged across simulated datasets
coverage	accuracy	percentage of simulated datasets in which the estimated 95% CI contained the true value
power	precision	percentage of simulated datasets in which the p-value for the estimate was less than a p-value threshold (0.05 used here)

Inputs

Data

An EMN plot contains three 45 m parallel transects. Plot locations were selected with a generalized random tessellation stratified (GRTS) survey design with modifications (see Appendix A). Shrub layer and ground layer plants are sampled in eight 1-square-meter quadrats per transect, spaced 5 m apart, and trees are sampled from a 45 m x 10 m subplot centered on each transect. Data collected include cover class of shrub/ground layer plants, amount of deer browse on woody species in the shrub layer, earthworm disturbance, and counts and diameter-at-breast-height (DBH) of trees. Deer browse is measured in a circle with a 1 m radius, centered on each quadrat, where any tree or shrub that intersects the circle is recorded and assessed for browsing. Detailed methods are described by Rowe (2022).

Indicator variables

We chose indicators for trends in statewide biodiversity that could be analyzed with the EMN dataset and that were included in a study that provided estimates needed for simulated datasets (Perles et al. 2014b) (Table 2). Note that sample size estimates for these indicators are applicable to indicators with the same form (e.g., proportion) and variance components. The richness indicators are formatted as proportions, to capture the relative contributions of native and non-native taxa, which may differ from count variables used in other contexts. We also performed analyses of trends in generic (undefined) univariate indicators at both the statewide and community scales. “Community-scale” refers to a subset of statewide plots that might be analyzed together, such as all plots sharing the same habitat-specific species pool (Zobel 2016, Keddy and Laughlin 2021). While we explored lower plot numbers in the statewide analyses, the estimated variance among plots was likely larger than what occurs within a community, so we divided plot and trend variance (see next section) by 10 to better simulate a community-scale indicator. We chose indicators for trends in community-scale biodiversity and impacts in forested communities using the literature (cited in Table 2) and the EMN project objectives (Rowe 2022). For indicators that required native/non-native designations, we used MNTaxa (MN DNR 2022). At the time of this analysis, the dataset contained 862 native taxa and 82 non-native taxa. For indicators that included cover, we used the midpoint of cover categories: < 2%, 2-10%, 10-25%, 25-50%, 50-75%, and 75-100%.

Variance components

Variance components are the contributions of different sources of variation (e.g., space, time) to total variance. We used variance components reported by Perles et al. (2014b) to build variation into the simulated datasets for many of the indicators. We used the variance components for native ground layer cover to simulate datasets for graminoid cover. We assumed generic indicators were proportions and had the same variance components and total variance as nonnative ground layer plant cover, which is a proportion with relatively high variance. The variance components for average cover of invasive exotic species (Perles et al. 2014b) were applied to taxa presence/absence for shifts in community composition. We assumed all variance for biotic homogenization metrics could be attributed to variance among plot pairs (estimated with existing EMN data) and residual variation (assumed value). We assumed only residual variance for the forest impact analyses because these analyses could be performed with only baseline data, limiting the influence of temporal variation. We confirmed this assumption by examining the subset of plots that were revisited. For final analysis of forest impacts, we

recommend including other plot-level measurements that could influence the response variable as model covariates to account for plot variation. True values for variance for each indicator were transformed based on the modelling approach (typically log-odds transformed).

Study designs

EMN had 316 existing plots, with the potential to collect more baseline data from additional plots over the next three field seasons. If 94-95 plots, which is equal to the maximum annual number sampled so far in the program, were sampled each year for the next three years, the total sample size would be 600 plots. EMN researchers plan to regularly revisit plots to estimate trends over time. We evaluated study designs that varied in plot number (sample size) and revisit frequency (once or four additional times). For statewide indicator values, we explored samples sizes of 316, 350, 400, 450, 500, 550, and 600. For the generic indicator values, we explored additional lower sample sizes. For the shift in community composition, we explored sample sizes equivalent to those in the dataset, double, and triple that number. For biotic homogenization, we explored sample sizes of 5, 10, 20, 30, and 40. For forest impacts, we explored samples sizes of 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, and 300. Designs with the number of plots that have already been completed assumed no further baseline data were collected while designs with more plots assumed an equal number of plots were sampled over the following three years. We assumed forest impact analyses were performed on a single cycle of data, so the number of revisits was not varied.

Table 2: Indicator values used in power analysis.

Scale	Driver	Indicator	Description
statewide	time	native ground layer plant richness	proportion of total species richness that belonged to native taxa, averaged across quadrats within a plot
statewide	time	native ground layer plant cover	proportion of summed cover that belonged to native taxa, averaged across quadrats within a plot
statewide	time	non-native ground layer plant richness	proportion of total species richness that belonged to non-native taxa, averaged across quadrats within a plot
statewide	time	non-native ground layer plant cover	proportion of summed cover that belonged to non-native taxa, averaged across quadrats within a plot
statewide	time	graminoid cover	proportion of summed cover that belonged to Poaceae, Juncaceae, or Cyperaceae families, averaged across quadrats within a plot
statewide/ community	time	generic	proportion an undefined indicator with variance components and total variance equal to non-native ground layer plant cover (variance among plots divided by 10 for community-scale)
community	time	shift in community composition	dissimilarity in ground layer plant composition (presence/absence) over time within a plot
community	time	internal biotic homogenization	dissimilarity in ground layer plant composition among plots within a community (Olden and Rooney 2006)

Scale	Driver	Indicator	Description
community	time	external biotic homogenization	dissimilarity in ground layer plant composition between two or more communities
forests	deer browsing	*Acer* occurrence	proportion of browse circles in which an *Acer* species is present (Frerker et al. 2013)
forests	earthworm disturbance	graminoid cover	see graminoid cover above (Fisichelli et al. 2013)
forests	timber harvesting	*Acer* basal area	π x (DBH/2)^2, averaged across *Acer* species (Bouchard and Pothier 2011)

Trends

Trend values represented the annual multiplicative change in the indicator:

$$y_j = y_0(1 + \lambda)^j,$$

where y_j is the indicator value at year j, y_0 is the indicator value at year 0, and λ is the trend value). We used annual trend values of -5% or +5%, which are large, but realistic (Perles et al. 2014b). A trend value of -/+5% amounts to an approximate change of -40% or +63%, respectively, over 10 years. We explored the effect of trend values on recommended sample sizes with the generic indicator. For shift in community composition, we assumed one-third of the taxa increased by 5%, one-third declined by 5%, and one-third did not change.

Baseline average and plot variance

The average of an indicator value across existing EMN plots served as the baseline average value in simulated datasets. We also estimated variance among plots using data from revisited EMN plots. Ten EMN plots were revisited after one year in 2018 (N = 9) or 2019 (N = 1). We fit mixed-effects models with the following formula to this data subset using the glmmTMB package (Brooks et al. 2017) in R (R Core Team 2024):

$$y \sim revisit + (1|plot),$$

where y is the indicator value, $revisit$ is a binary variable indicating whether the observation was from the original visit or the revisit, and $(1|plot)$ estimates the desired among-plot variance. For shift in community composition, we fit this model, with an additional random intercept and slope for each taxon, to data from taxa that occurred in 3-9 plots during each of the original visits and revisits.

Forest impacts effect sizes

For forest impacts, we assumed changes in the indicator variables (Table 2) were linearly related to changes in browse pressure, earthworm disturbance, and timber harvesting. We used EMN data from forest plots (fire-dependent forests, mesic hardwood forests, floodplain forests, wet forests, and forested peatlands) (MN DNR 2003, 2005a, 2005b) to estimate the effects of deer browse pressure and earthworm disturbance on *Acer* occurrence and graminoid cover, respectively. We did not have data on timber harvest history of plots, so we assumed effect sizes equal to those explored for deer browse pressure and earthworm disturbance.

Deer browse pressure was quantified as the number of browsed species, summed across all deer-browse quadrats, divided by the occurrence of all species, summed across deer-browse quadrats (Frerker et al. 2013). We estimated the effect of deer browse pressure on abundance of the most common genus in deer-browse quadrats by fitting a linear regression to the log-odds transformation of the proportion of quadrats *Acer* species occupied with deer browse pressure as the predictor variable.

To evaluate an effect of earthworms on the ground layer plant community, we fit a linear regression to log-odds transformed graminoid cover with earthworm disturbance as the predictor variable (Fisichelli et al. 2013). Earthworm disturbance is an ordinal variable ranging from one to five that captures multiple indicators of earthworm activity (Loss et al. 2013). We used the average of two earthworm disturbance values per plot.

From these models, we used the coefficient estimate as one of our effect inputs, the intercept as our baseline average, and the residual standard deviation as the residual standard deviation in our simulated datasets. We also explored effect sizes slightly greater and slightly less than the observed effect.

Simulating data

Trends

To simulate trends in univariate indicators, we used a model of long-term trends that takes into account variation from multiple sources (Urquhart 2012, Perles et al. 2014b):

$$y_{ij} = \mu + a_i + year_j(\lambda + t_i) + b_j + e_{ij},$$

where y_{ij} is the value of the indicator in plot i and year j, μ is the baseline average, a_i is among-plot variance, $year_j$ is number of years elapsed from the first year, λ is the overall trend, t_i is plot-specific deviation from the overall trend, b_j is among-year variance, and e_{ij} is residual variance. We used EMN data and trend values described above for μ and λ, respectively, for each indicator. All indicators assessed were proportions and we used the log-odds (or logit) transformation, where $y_{ij} = ln[q/(1-q)]$ for proportion q. Because all ground layer plants were categorized as native/non-native and to family, the indicators can be interpreted as the natural log of the ratio of plants (richness or cover) in the category to those not in the category (e.g., *ln(native/non-native)*).

We drew variance values from normal distributions with a mean value of zero and standard deviations informed by among-plot variation in EMN data and variance components from Perles et al. (2014b). Variance components were informed by data collected in Delaware Gap National Recreation Area, which had approximately 128 observations between 2007-2011 (Perles et al. 2014b). The dataset included one set of plots sampled in the first and last year and one set of annually sampled plots. The annually sampled plots were sampled twice per year, within the span of a few weeks (Perles et al. 2014a, 2014b). Perles et al. (2014b) reported the percentage of variance attributed to each source in the above equation, which we used to parameterize t_i, b_j, and e_{ij}. We used the plot variance estimate from EMN, described above, as a_i.

Community composition shifts

We used the top three most commonly sampled community types in the existing EMN data, which were defined by unique combinations of Native Plant Community system and floristic region within Ecological Classification System sections (MN DNR 2003, 2005a, 2005b): southern upland prairie (UPs) in the North Central Glaciated Plains, northern fire-dependent forests (FDn) in the Northern Superior Uplands, and southern mesic hardwood forests (MHs) in the Paleozoic Plateau. To simulate shifts in ground layer plant community composition, we started by simulating the presence of each taxon within the community using a model of long-term trends that allows for correlations among groups of taxa (Irvine et al. 2011, Urquhart 2012, Perles et al. 2014b):

$$y_{ijl} = s_l + a_i + year_j(\lambda_l + t_i) + b_j + c_{ij} + e_{ijl},$$

where y_{ijl} is the log-odds transformed probability taxon l is present in plot i and year j, s_l is the taxon-specific intercept (average across all plots), a_i is among-plot variance, $year_j$ is number of years elapsed from the first year, λ_l is the taxon-specific trend (increasing/decreasing across all plots), t_i is plot-specific deviation from the overall trend, b_j is among-year variance, c_{ij} is plot-year interaction variance, and e_{ijl} is residual variance. This probability was used to assign each taxon a presence/absence value for each plot in each year.

To incorporate concurrent trends for groups of taxa, Irvine et al. (2011) drew the parameters s_l and λ_l from a multivariate normal distribution with mean and variances assigned to groups rather than taxa. That is, all taxa within a group have an average frequency (or cover) and trend that is correlated with other taxa in the group. We simplified this step by using the log-odds transformed average presence of each taxon over all plots as the taxon-specific intercept and we randomly assigned taxa one of three trend values: increasing, decreasing, or no change.

Biotic homogenization

To simulate biotic homogenization, we assumed dissimilarity changed between monitoring cycles, with variation among plot pairs described as:

$$d_{pc} = \nu + g_p + cycle_c\gamma + e_{pc},$$

where d_{pc} is the log-odds transformed dissimilarity between plot pair p in cycle c, ν is the baseline average, g_p is the variance among plot pairs, $cycle_c$ is the number of monitoring cycles (or revisits) since the baseline visit, γ is the change in dissimilarity, and e_{pc} is the residual variance. To get the baseline average, we took the mean of Jaccard dissimilarity indices for all relevant pairwise comparisons in the EMN dataset (i.e., all plots within a community type for internal biotic homogenization and all comparisons between plots of different community types for external biotic homogenization). For internal biotic homogenization, we used the same community types as those used for composition shifts. For external biotic homogenization, we used the following pairs of community types: northern and southern upland prairie (UPn and UPs), central and northern fire-dependent forest (FDc and FDn), and central and southern mesic hardwood forest (MHc and MHs) (MN DNR 2003, 2005a, 2005b). We used the standard deviation of these dissimilarity indices as the standard deviation for g_p, which follows a normal distribution with a mean of zero.

Forest impacts

To simulate forest impacts, we used the effect sizes, baseline averages, and residual standard deviations from the fitted linear regressions for deer browse and earthworms described in *Forest impacts effect sizes*. To generate a range of values for the predictor variables, we randomly sampled, with replacement, from the deer browse pressure and earthworm disturbance values measured in EMN forest plots. For timber harvesting, we calculated the baseline average and standard deviation (used as residual standard deviation) of *Acer* basal area by averaging the DBH of all *Acer* trees measured in a plot, then converting to basal area with the formula described in Table 2. We used years since harvest as the predictor variable, which was a random variable drawn from a Poisson distribution with an average of ten.

Fitting models

Trends

To evaluate study designs, we fit statistical models to simulated datasets. For trends in univariate indicators, we estimated the overall trend, spatial variance, trend variance, and temporal variance with the following formula using the R package glmmTMB (Brooks et al. 2017):

$$y \sim wyear + (1 + wyear|plot) + (1|year),$$

where y is a simulated observation and $wyear$ is the number of years since the start of the study. The term $(1 + wyear|plot)$ estimated variation among plots that is time-invariant (i.e., plot random intercept) and variation among plots in the overall trend (i.e., plot random slope). For indicator values with low trend variance (< 0.001), we removed the random slope. The term $(1|year)$ estimated variation among years (i.e., year random intercept). We log-odds transformed proportion data and used a normal mixed-effects regression. We ran several analyses with binomial models for proportion data, and trend estimates seemed to be biased positive for indicators with high baseline values and negative for indicators with low baseline values. More simulations would be needed to determine the cause of this directional bias.

Community composition shifts

To evaluate shifts in plant community composition, we used constrained correspondence analysis (CCA) and distance-based redundancy analysis (dbRDA) with the *cca* and *dbrda* functions, respectively, in the R package vegan (Oksanen et al. 2022). Both methods have been used on presence/absence data (Irvine et al. 2011, van der Veen et al. 2023) and differ in their calculations of community dissimilarity and regression methods (Oksanen et al. 2022). We used the binary formula of the Bray-Curtis dissimilarity metric for dbRDA. For both methods, the predictive formula included $wyear$ as a constraint and $plot$ as a conditioned variable, which was partialled out before the constraint was assessed. We then conducted a permutation test with the function *anova* to calculate a p-value for $wyear$. Permutations were restricted to among plots but did not occur within plots. We did not evaluate measures of precision and accuracy other than power because separate trend values were input for each taxon, but the output of this analysis is at the community scale. For analyses of EMN data in the future, we recommend joint species distribution models to evaluate community composition (Appendix B).

Biotic homogenization

To evaluate biotic homogenization, we fit normal linear mixed-effects models with the formula

$$d \sim cycle + (1|pair),$$

where d is a simulated dissimilarity, $cycle$ is the number of cycles since the start of the study, and $(1|pair)$ is the plot-pair intercept (estimated variation among plot pairs).

Forest impacts

Forest impacts were assessed with normal linear regressions and the response and predictor variables described in Table 2, *Forest impacts effect sizes*, and *Simulating data: Forest impacts*. The response variables of deer browse and earthworms were proportions and were log-odds transformed while the response variable of timber harvesting was non-negative continuous and log-transformed

Analyses with random and non-random plots

The final EMN dataset will have plots that were selected semi-randomly and non-randomly (Appendix A), which can be included in the same analyses (Lookingbill et al. 2012, Irvine et al. 2018). Plots are assigned design weights to account for differences in inclusion probability. In simple designs, the design weight is the inverse of the inclusion probability. An example of a simple inclusion probability is $\frac{n}{N}$, where n is the area covered by the samples in the study and N is the area covered by the population. Irvine et al. (2018) assume that legacy sites–older sites included in the sample in order to collect revisit data–have weights of one, indicating that inference from these sites only applies to the site and not a population. For plots that are part of a GRTS survey design, but selected through a non-random process, Lookingbill et al. (2012) suggests that the inclusion probability is a combination of the probability the plot was selected with GRTS and the probability the plot was selected through the non-random process. It is necessary to define the non-random process to estimate this probability.

We performed analyses of native ground layer plant richness and graminoid cover assuming a range of sample sizes for non-randomly selected plots. We assigned the design weight for a plot chosen with the GRTS algorithm as:

$$(Nn_T)/(n_E n_R),$$

where N is the size of the target population (e.g., the area of Minnesota), n_T is the summed area of GRTS points (assuming each point is the size of an EMN plot) that are within the target population, n_E is the summed area of GRTS points that were evaluated, and n_R is the summed area of EMN plots in which the response was measured (Starcevich et al. 2016). Note that this design weight formula is the inverse of multiple inclusion probabilities multiplied together: the probability that a GRTS point is evaluated and the probability that a target point is visited. Another way to think about this is that we are adjusting the size of N based on our knowledge. For example, if only 20% of the evaluated plots are target plots, we multiply N by 20% and then divide by the number of sampled plots to get the design weight, or inverse of the inclusion probability. We assigned the design weight of a plot chosen for a specific purpose as 1 (Irvine et al. 2018). We represent this by:

$$n_N/n_N \ ,$$

where n_N is the summed area of purposely chosen plots. The design weight of a plot associated with a GRTS point, but not chosen with the algorithm (i.e., out of order, targeting a specific area or plant community) should be between one and the design weight for plots chosen with the GRTS algorithm. We do not yet know the best way to calculate these design weights. To combine plots with different design weights together in an analysis, each design weight is divided by the sum of the design weights (Nahorniak et al. 2015). The purpose of adjusting the design weights is to reduce the bias caused by including the non-randomly chosen plots (Irvine et al. 2018).

Results

Statewide analyses

Please note that throughout the results, the y-axis on each figure is calibrated to the values within the specific figure (i.e., the y-axis differs among figures). Baseline means are provided in tables on the original, not transformed, scale that is described in Table 2.

Native ground layer plant richness

Average native ground layer plant richness was relatively high in the baseline dataset, and we expect there to be high variation in this indicator among plots (Table 3). The averages of models fit to simulated datasets indicated high precision and accuracy across study designs and trend values (Fig. 1A). There was some variation in bias across study designs and trend values, but estimates were on average within 0.4% of the input trend (Fig. 1B). More plot revisits increased precision, indicated by smaller 95% CI (Fig. 1A, C), and accuracy, indicated by higher coverage (Fig. 1D). Power was 100% for all study designs with four revisits (Fig. 1E). For study designs with one revisit, increasing the number of plots from 316 to 350 increased power by about 15% and increasing the sample size further increased power by at most 5% (Fig. 1E).

Table 3: Simulated data inputs of statewide native ground layer plant richness analysis.

Input	Value
baseline mean	0.9378
spatial (plot) variance (a_i)	4.756
temporal (years) variance (b_j)	0.1191
trend (by plot) variance (t_i)	0
residual variance (e_{ij})	0.5622
completed plots	316
non-random trend	0

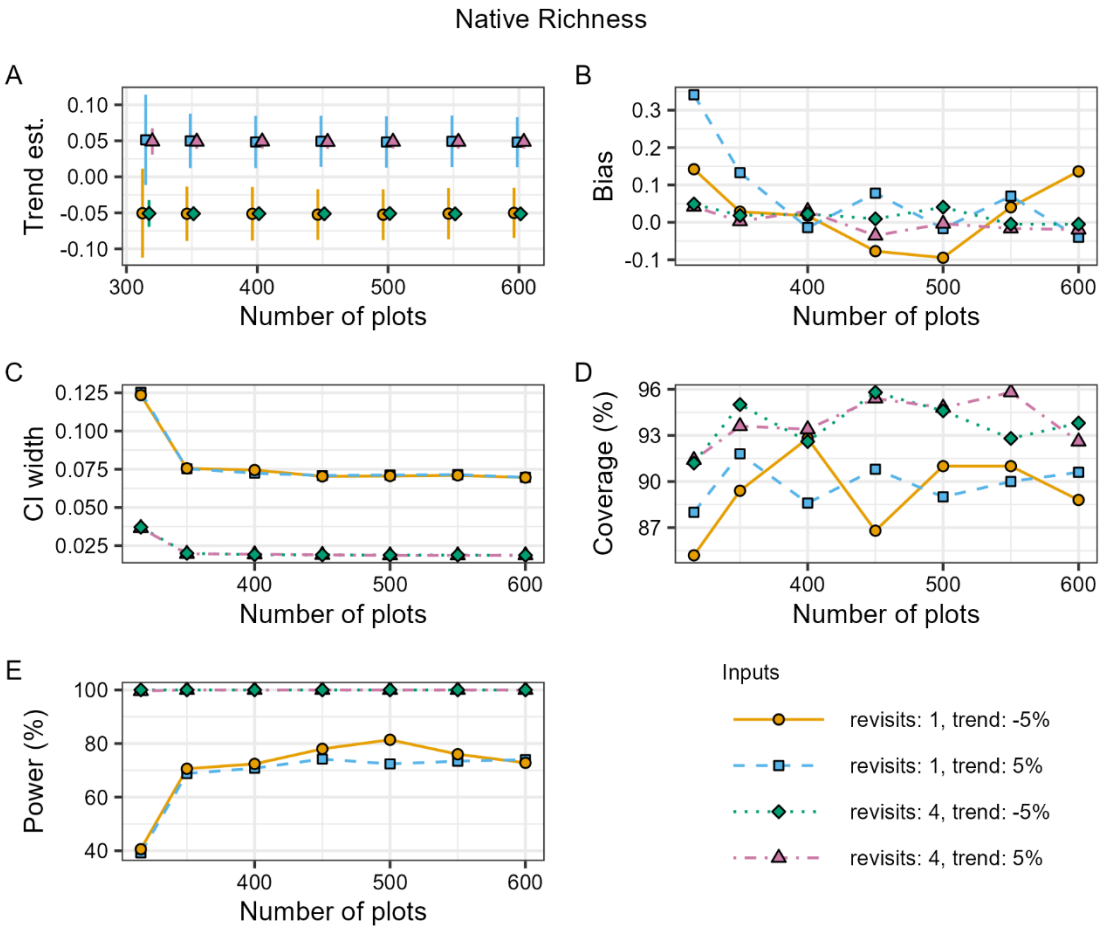


Figure 1: Power analysis of native ground layer plant richness. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

When the total sample size was set to 600, the proportion of plots that were selected non-randomly did not meaningfully influence the accuracy and precision of native ground layer richness estimates (Fig. 2A). Note that the assumed trend of randomly chosen plots (-5%) differed from that of non-randomly chosen plots (0%, Table 3). We would expect greater accuracy with more similar trend values between the two plot types. Bias was within 0.04% of the input trend (Fig. 2B). Precision slightly increased when study designs had more random plots, but the precision was generally high across all study designs (Fig. 2C). Similarly, coverage exceeded 90% for all study designs and did not change monotonically with the number of random plots (Fig. 2D). Power was 100% across all study designs (Fig. 2E).

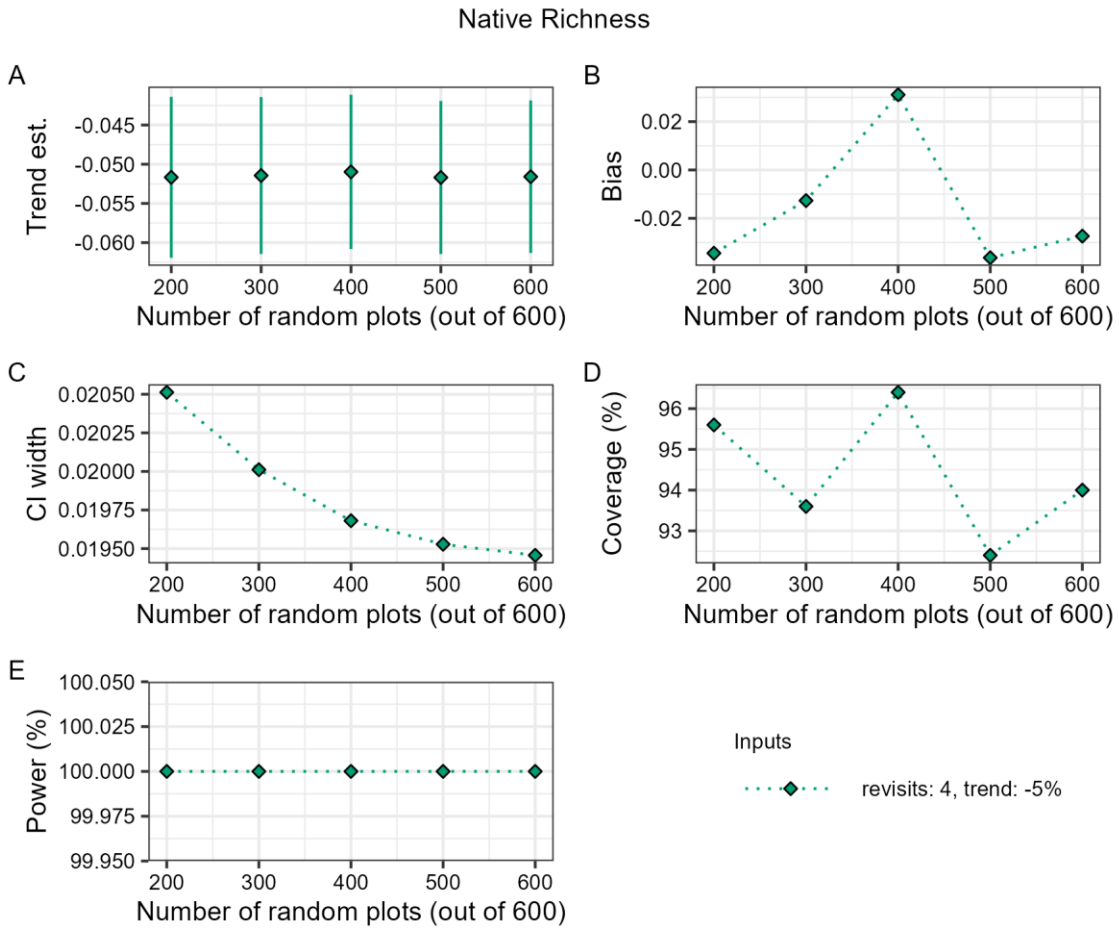


Figure 2: Power analysis of native ground layer plant richness with incremented inclusion of non-random plots (out of 600). (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$. In B-E, simulations are represented by triangles connected by dashed lines.

Native ground layer plant cover

Native ground layer plant cover (Table 2) had greater assumed spatial and trend variances than native ground layer plant richness (Table 4). Because the two indicators were both proportions with similar baseline averages and the same trend values, we can roughly compare the results. Despite greater spatial and trend variances, the accuracy and precision in trend assessments for native ground layer plant cover remained relatively high (Fig. 3A). Trend estimates were within 0.15% of the input trend values for sample sizes less than 400 and within 0.1% for sample sizes greater than 400 (Fig. 3B). CI width declined with increasing sample sizes and more revisits (Fig. 3C). Coverage exceeded 93% across all study designs and converged around 95% for study designs with 600 plots, regardless of revisit frequency or the direction of the trend (Fig. 3D). Power increased from approximately 85% to 100% with increasing sample sizes for study designs with four revisits (Fig. 3E). For study designs with one revisit, power exceeded 80% with at least 350 plots (Fig. 3E).

Table 4: Simulated data inputs of statewide native ground layer plant cover analysis.

Input	Value
baseline mean	0.9182
spatial (plot) variance (a_i)	6.6695
temporal (years) variance (b_j)	0
trend (by plot) variance (t_i)	0.0812
residual variance (e_{ij})	0.5655
completed plots	316

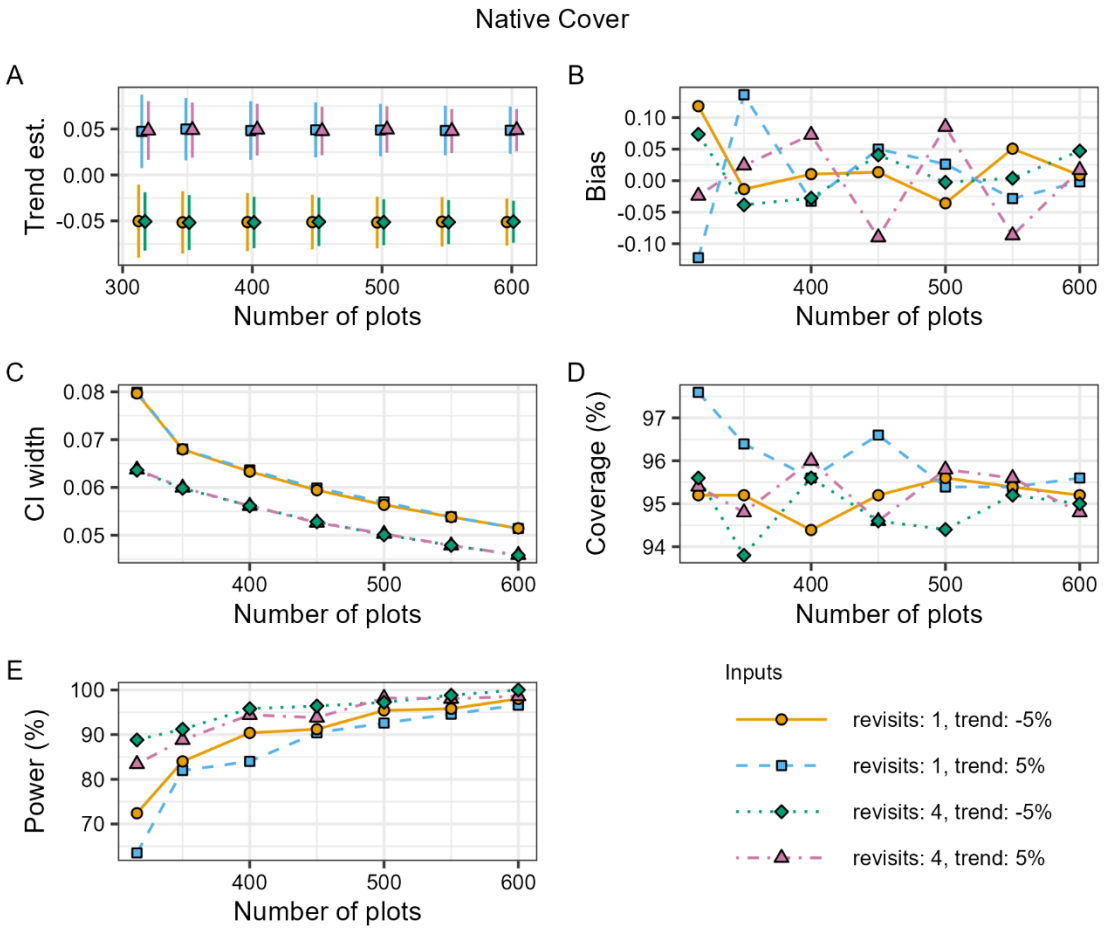


Figure 3: Power analysis of native ground layer plant cover. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Non-native ground layer plant richness

Non-native ground layer plant richness (Table 5) had a much lower baseline average than the native ground layer indicators (Table 3), but similar variance values. Precision and accuracy of the models for non-native ground layer plant richness were generally higher than those for native ground layer plant richness, potentially due to lower residual and temporal variation, but otherwise displayed similar patterns (Fig. 4). Specifically, increasing the sample size from 316 to 350 plots in study designs with one revisit led to the greatest increase in precision (Fig. 4C, E). Trend estimates were within 0.1% of input trend values across all study designs (Fig. 4B) and power exceeded 95% for all study designs with at least 350 plots (Fig. 4E).

Table 5: Simulated data inputs of statewide non-native ground layer plant richness analysis.

Input	Value
baseline mean	0.0622
spatial (plot) variance (a_i)	4.756
temporal (years) variance (b_j)	0.007
trend (by plot) variance (t_i)	0.0488
residual variance (e_{ij})	0.1719
completed plots	316

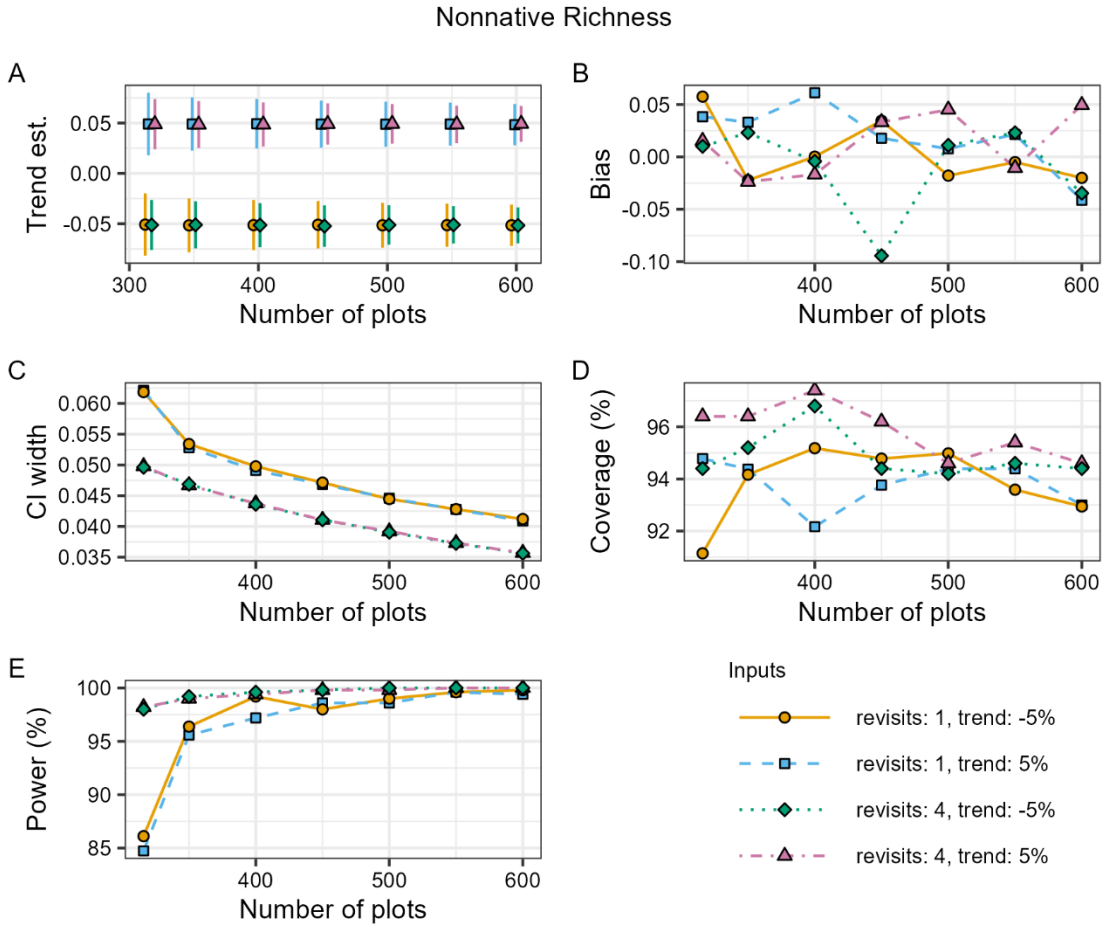


Figure 4: Power analysis of non-native ground layer plant richness. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Non-native ground layer plant cover

Non-native ground layer plant cover had the highest trend variation of the statewide biodiversity indicators (Table 6). Accuracy was slightly lower than other indicators, with trend estimates within 0.3% of true trend values and coverage greater than 90% for all study designs (Fig. 5B, D). The confidence intervals around trend estimates were relatively large (Fig. 4A, C), leading to a maximum power value of approximately 55% (Fig. 4E). Power increased with sample size and revisits (Fig. 4D).

Table 6: Simulated data inputs of statewide non-native ground layer plant cover analysis.

Input	Value
baseline mean	0.0818
spatial (plot) variance (a_i)	6.6695
temporal (years) variance (b_j)	0.0468
trend (by plot) variance (t_i)	0.3799
residual variance (e_{ij})	0.2233
completed plots	316

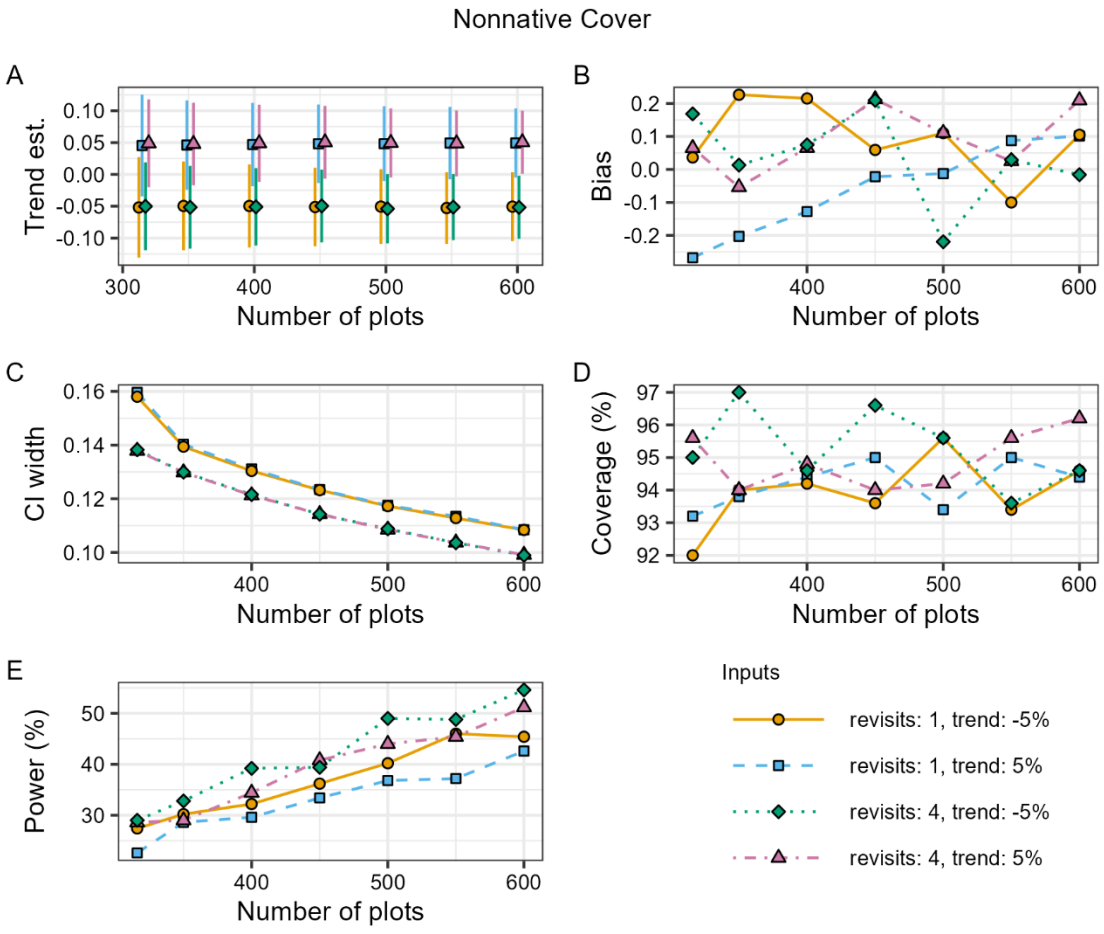


Figure 5: Power analysis of non-native ground layer plant cover. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Graminoid cover

Models of graminoid cover, which features a functional group of ground layer plants rather than a group based on origin, had similar accuracy and precision to those for native cover, which supplied the variance components (Table 7, Fig. 6). All study designs have high accuracy, with trend estimates within 0.15% of input trends (Fig. 6B) and coverage greater than 93% (Fig. 6D). Four revisits of any sample size or one revisit with at least 400 plots led to power greater than 90% (Fig. 6E).

Table 7: Simulated data inputs of statewide ground layer graminoid cover analysis.

Input	Value
baseline mean	0.2704
spatial (plot) variance (a_i)	5.8268
temporal (years) variance (b_j)	0
trend (by plot) variance (t_i)	0.0709
residual variance (e_{ij})	0.4941
completed plots	316
non-random trend	0

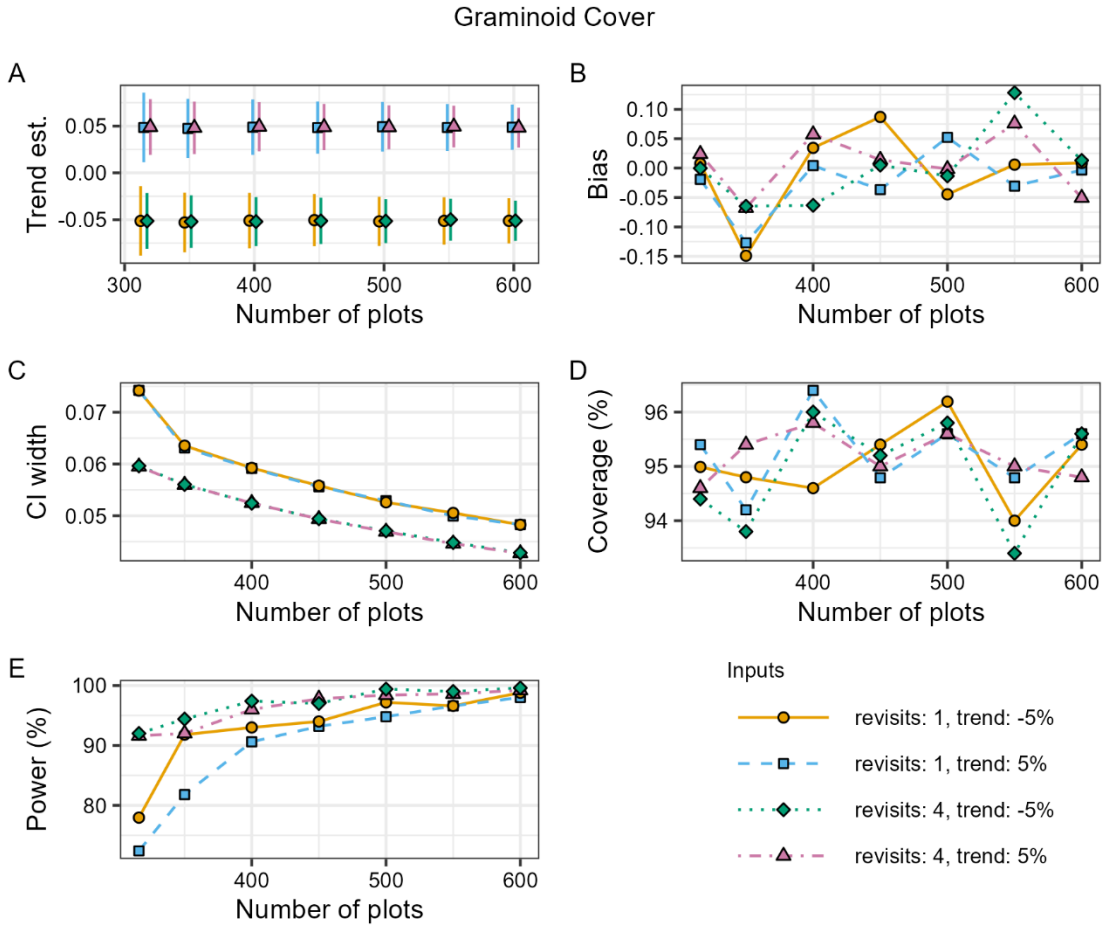


Figure 6: Power analysis of graminoid cover. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

When the total sample size was set to 600, the proportion of plots that were selected non-randomly reduced precision (Fig. 7A, C, E). However, power exceeded 92% for all study designs (Fig. 7E). Trend estimates were within 0.2% of the trend input (Fig. 7B) and coverage exceeded 70% (Fig. 7D) for all study designs. These measures of accuracy generally improved as the proportion on non-random plots declined below 50% (i.e., 300 or more random plots).

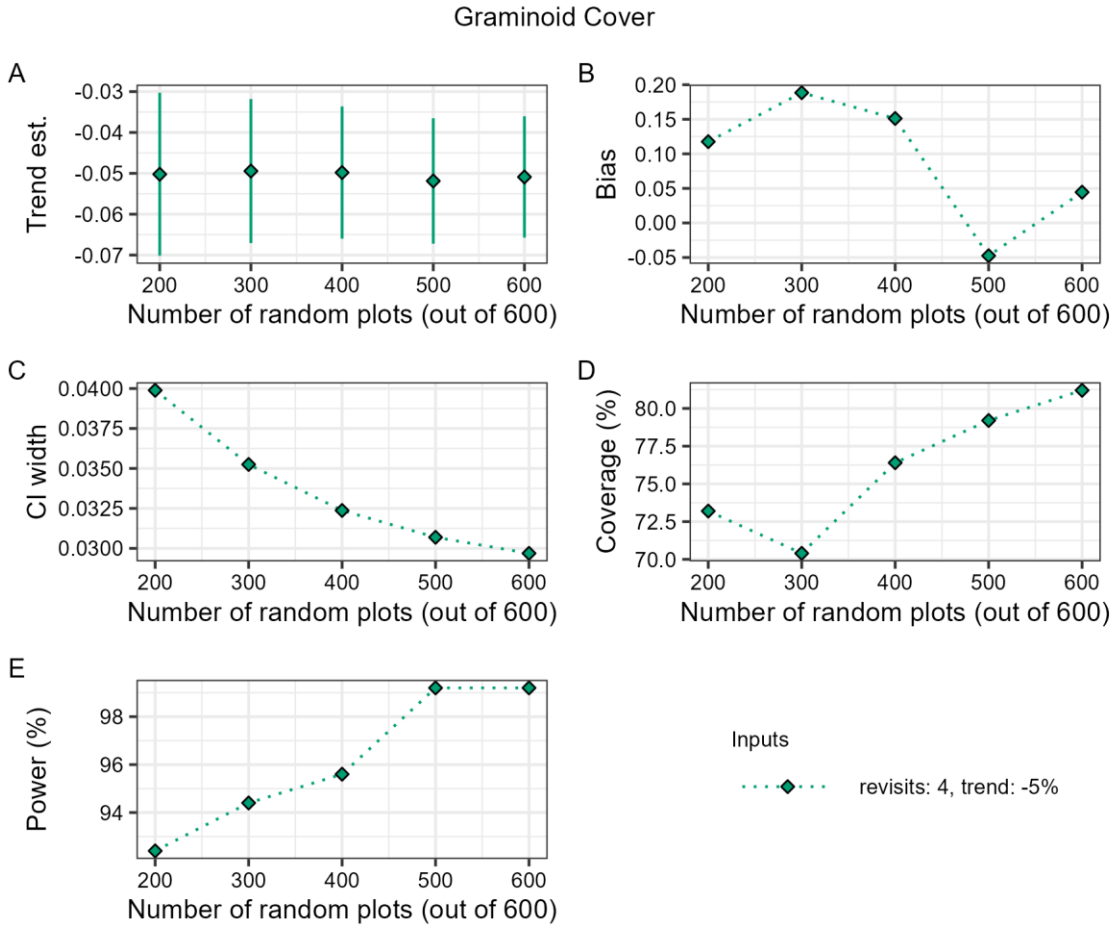


Figure 7: Power analysis of graminoid cover with incremented inclusion of non-random plots (out of 600). (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$. In B-E, simulations are represented by triangles connected by dashed lines.

Generic indicator

Models of a generic indicator suggest that higher trend values can increase power when trend variation is high (as was the case with non-native ground layer plant cover, which supplied the variance components) (Fig. 8E). Assuming four revisits and a baseline average of 0.5, power reached approximately 87% with at least 400 plots (Fig. 8E). Trend estimates were within 1% of true trend values when sample sizes were at least 100 plots (Fig. 8A), suggesting that lower sample sizes may result in inaccurate trend estimates given comparable variance values (Table 8). Precision, as measured through CI width, increased with sample size, achieving less than 0.2 (approximately twice the magnitude of the largest trend inputs) with at least 200 plots (Fig. 8C). Because of these large CI's, coverage was relatively high across study designs (Fig. 8D). We found the same results when we increased and decreased the baseline average values (Appendix C).

Table 8: Simulated data inputs of statewide generic indicator analysis.

Input	Value
baseline mean	0.5 (Fig. 8), 0.05 and 0.95 (Appendix C)
spatial (plot) variance (a_i)	6.6695
temporal (years) variance (b_j)	0.0468
trend (by plot) variance (t_i)	0.3799
residual variance (e_{ij})	0.2233
completed plots	0
revisits	4

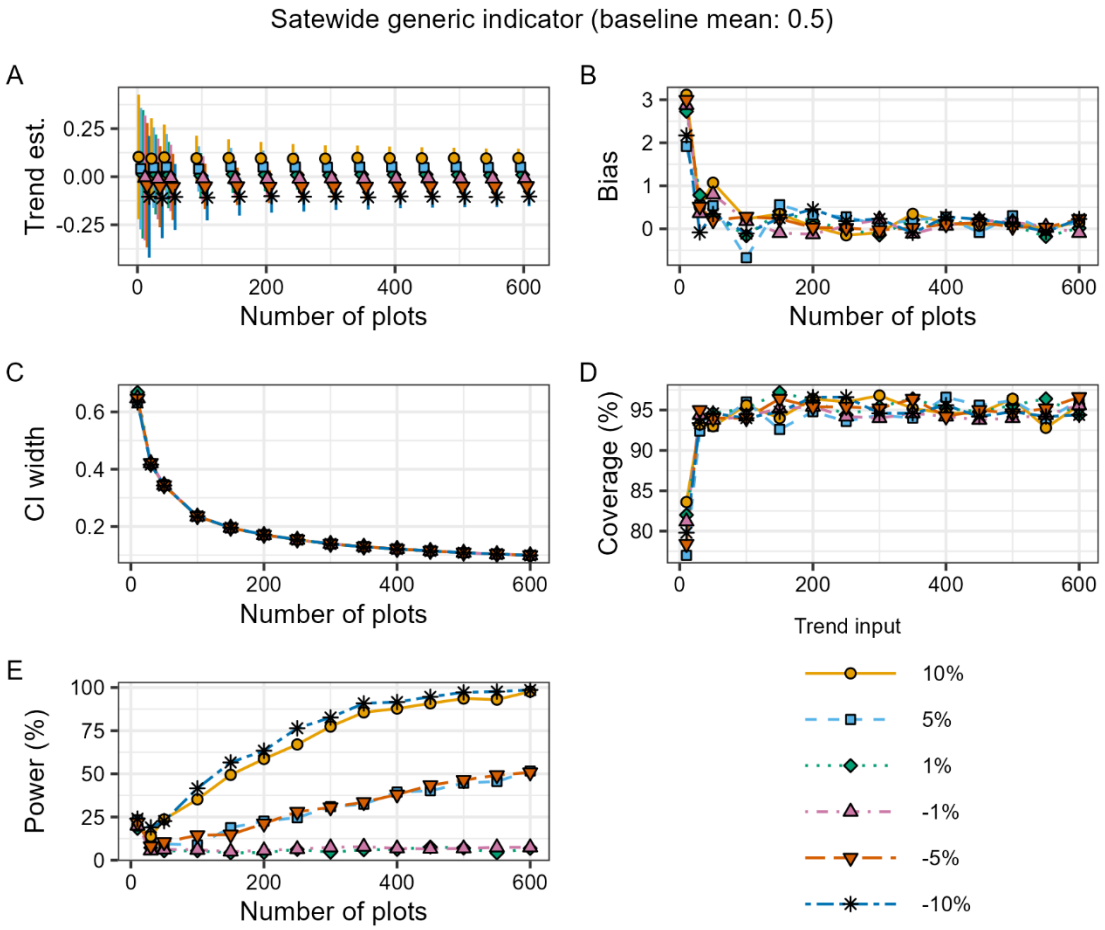


Figure 8: Power analysis of statewide generic indicator. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of

95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with P < 0.05.

Community analyses

Generic indicator

Trend estimates were within 0.6% of true trend values for sample sizes less than 100 plots and within 0.2% for sample sizes of at least 100 plots (Fig. 9B). Precision improved with sample sizes and CIs were less than 0.1 (approximately equal to greatest magnitude of trend input) for sample sizes greater than 100 plots (Fig. 9C). The greatest gains in coverage were between 10 and 30 plots, exceeding 90% for study designs with at least 30 plots (Fig. 9D). Power exceeded approximately 87% with at least 50 plots and trends of +/-10% or 150 plots and trends of +/-5% (Fig. 9E). We found the same results when we increased and decreased the baseline average values (Appendix C).

Table 9: Simulated data inputs of community-scale generic indicator analysis.

Input	Value
baseline mean	0.5 (Fig. 9), 0.05 and 0.95 (Appendix C)
spatial (plot) variance (a_i)	0.6669
temporal (years) variance (b_j)	0.0468
trend (by plot) variance (t_i)	0.038
residual variance (e_{ij})	0.2233
completed plots	0
revisits	4
non-random trend	0

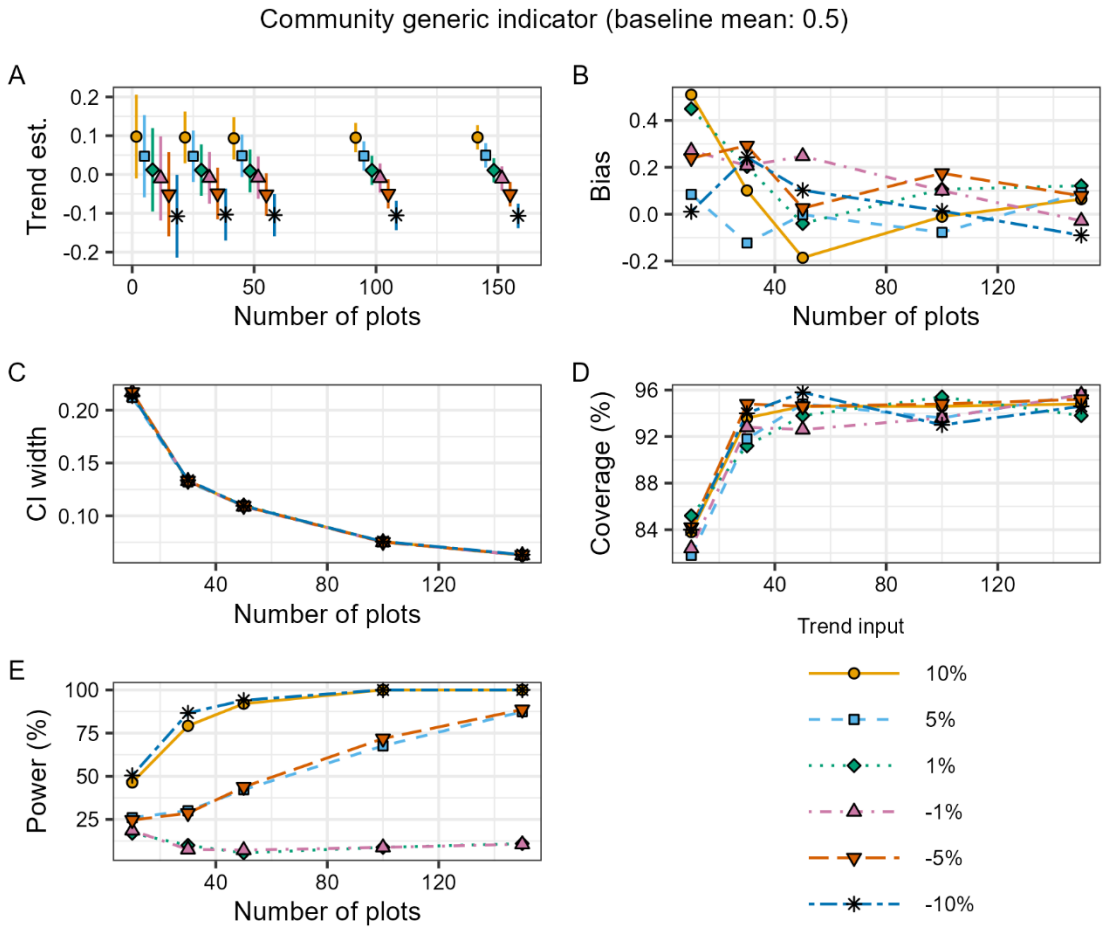


Figure 9: Power analysis of community-scale generic indicator. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

When we held the sample size at 100 plots and varied the number that were selected randomly, precision generally increased with more random plots while accuracy did not change monotonically (Fig. 10). More non-randomly selected plots did not increase bias or decrease coverage substantially relative to the study design with 100% randomly selected plots (Fig. 10B, D). The greatest decline in CI width (increase in precision) occurred between 10% and 25% of plots chosen randomly (Fig. 10C). Power increased nearly linearly with increasing proportion of randomly selected plots (Fig. 10E).

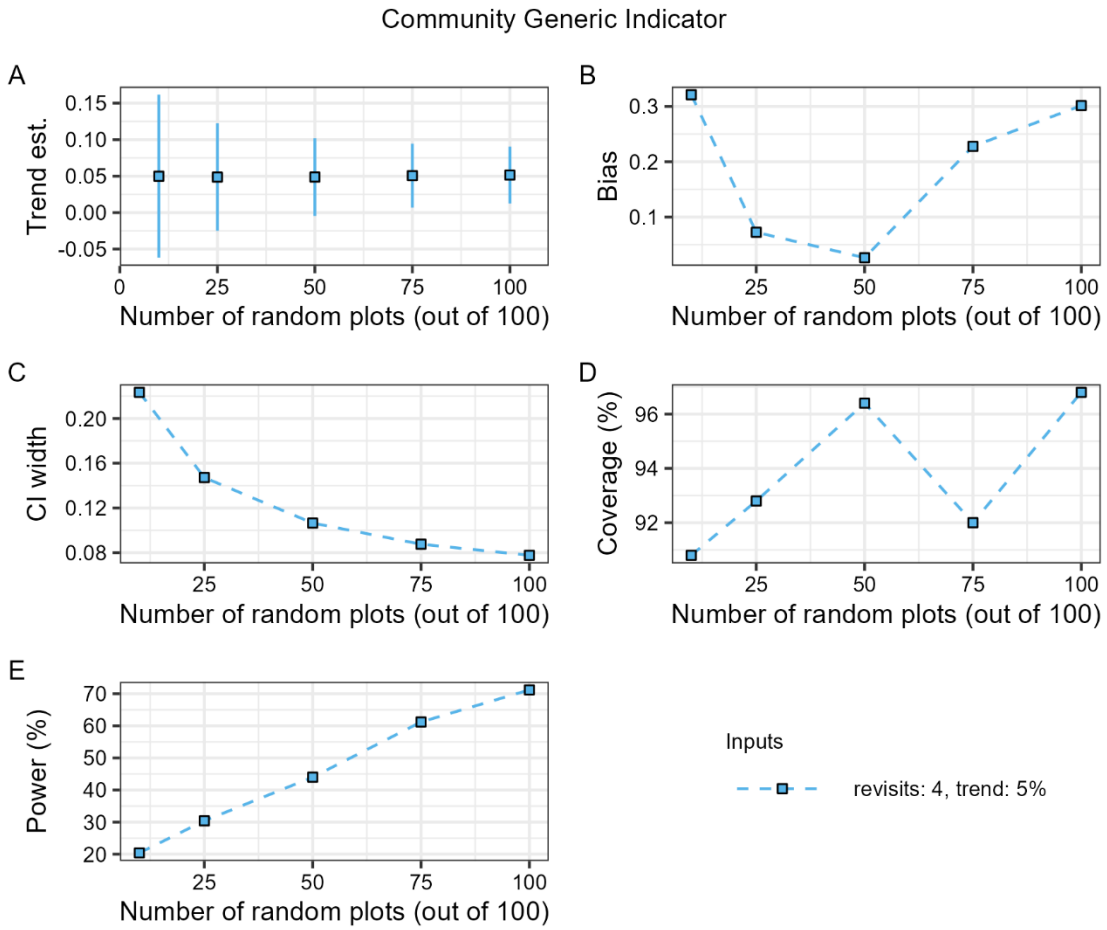


Figure 10: Power analysis of community-scale generic indicator with incremented inclusion of non-random plots (out of 100). (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Community composition

Baseline means in Table 10 are the proportion of plots in which taxa occurred, averaged across taxa within the community. Models using CCA could detect shifts in community composition with at least 87% power for study designs with at least 30 plots and four revisits or at least 90 plots and one revisit (Fig. 11). North Central Glaciated Plains UPs had almost equivalent sample size and variance components to Northern Superior Uplands FDn, but almost 20% lower power with the existing number of plots and one revisit (Fig. 11), indicating that communities with lower baseline averages (more rare species or fewer common species) may require more plots for similar power. Models using dbRDA required greater samples sizes than those using CCA for the same power (Fig. 12).

Table 10: Simulated data inputs for binary community composition analysis.

Input	North Central Glaciated Plains UPs	Northern Superior Uplands FDn	Paleozoic Plateau MHs
baseline mean	0.1781	0.2057	0.2136
spatial (plot) variance (a_i)	0.9997	0.9997	0.9997
temporal (years) variance (b_j)	0	0	0
trend (by plot) variance (t_i)	0	0	0
plot-time interaction variance (c_{ij})	0.0054	0.0054	0.0054
residual variance (e_{ijl})	0.0129	0.0129	0.0129
completed plots	31	30	24
trends	-5%/0%/5%	-5%/0%/5%	-5%/0%/5%

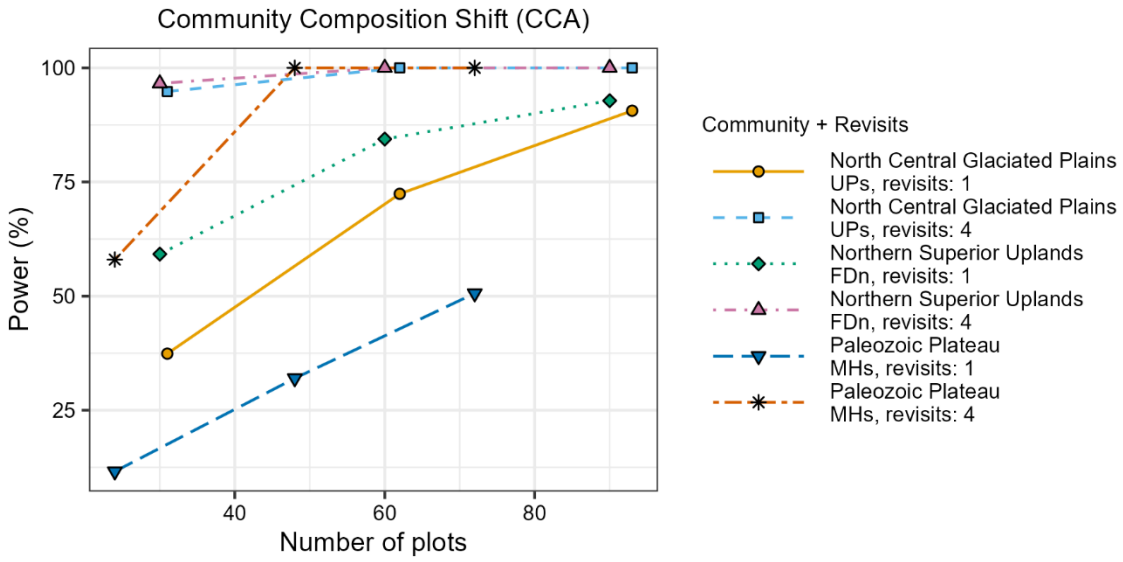


Figure 11: Power analysis of binary community composition using CCA (percentage of models with P < 0.05) for three communities, each with two revisit intervals.

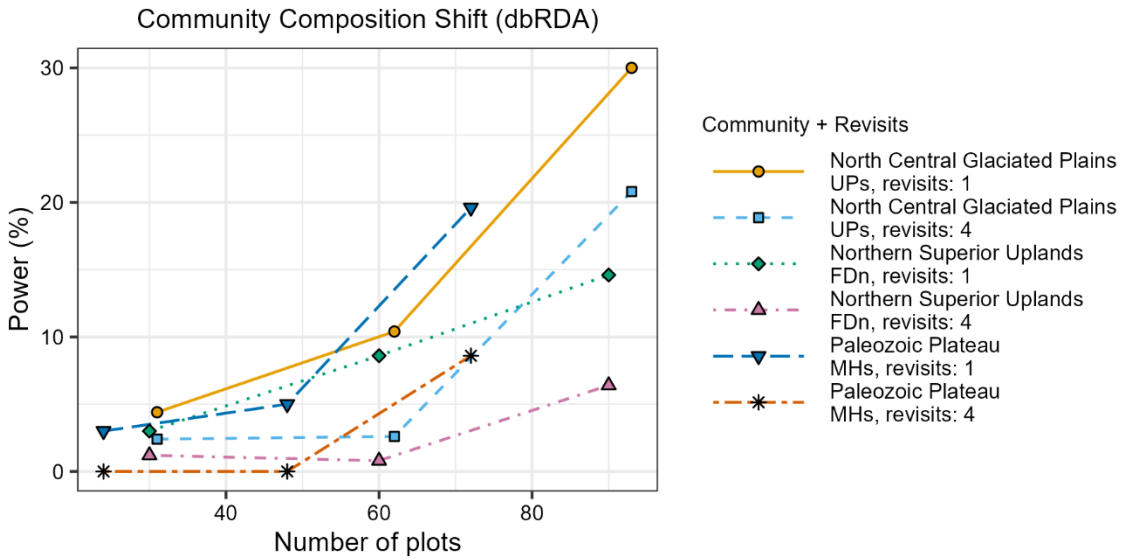


Figure 12: Power analysis of binary community composition using dbRDA (percentage of models with P < 0.05) for three communities, each with two revisit intervals.

Internal biotic homogenization

To evaluate internal biotic homogenization, we used the same communities evaluated for shifts in community composition. Averages and variance of dissimilarity values were similar for the three communities (Table 11). Precision and accuracy increased with sample size (Fig. 13A). For study designs with one revisit, accuracy and precision stabilized at their greatest value around approximately 20 plots (Fig. 13B-E). When study designs had four revisits, accuracy and precision were high for sample sizes as low as five plots (Fig. 13B-E).

Table 11: Simulated data inputs of internal biotic homogenization analysis.

Input	North Central Glaciated Plains UPs	Northern Superior Uplands FDn	Paleozoic Plateau MHs
baseline mean	0.7617	0.651	0.7569
plot pair variance (g_p)	0.0054	0.0095	0.0054
residual variance (e_{pc})	0.01	0.01	0.01
completed plots	31	30	24
trend (cycle)	-5%	-5%	-5%

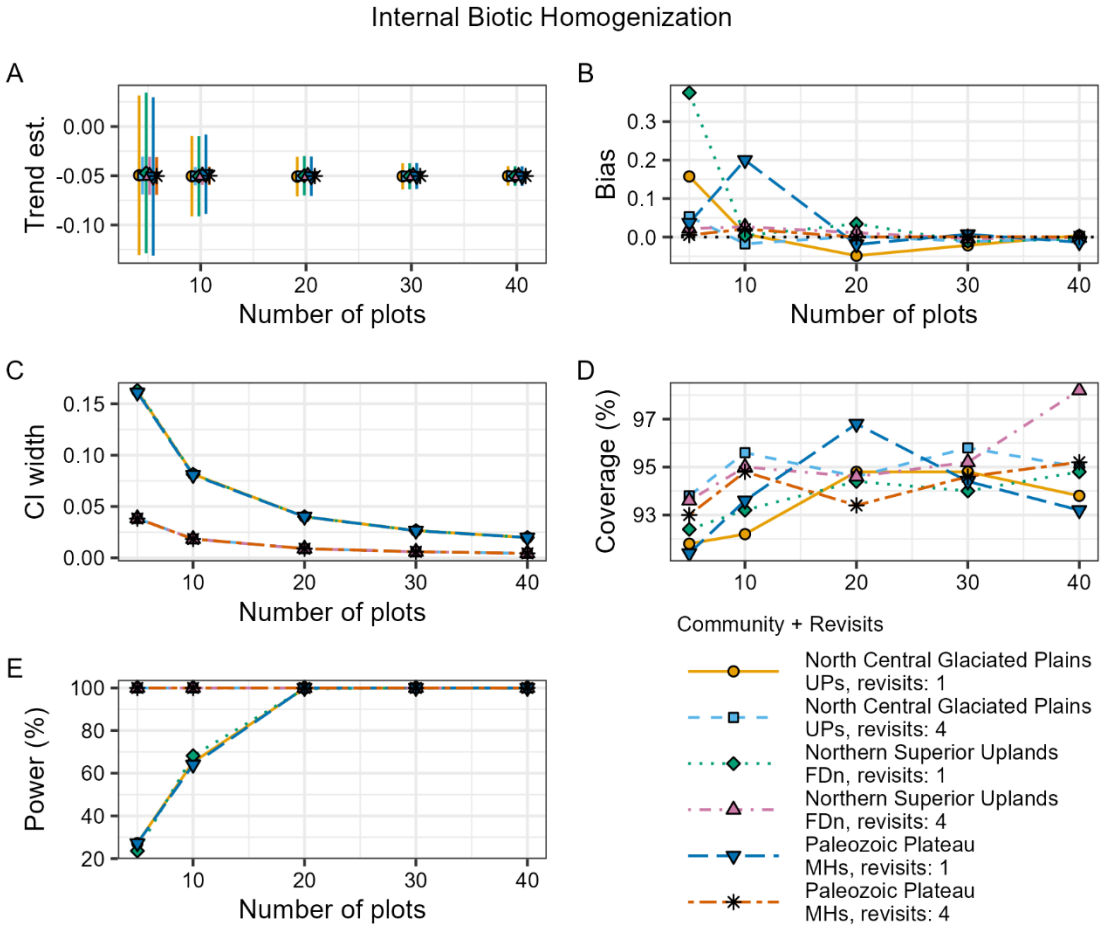


Figure 13: Power analysis of internal biotic homogenization. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

External biotic homogenization

To evaluate external biotic homogenization, we used the same NPCs evaluated for shifts in community composition, paired with an NPC in a neighboring floristic region. Averages and variance of dissimilarity values were similar for the three community pairs (Table 12). Precision and accuracy increased with sample size (Fig. 14A). For study designs with one revisit, accuracy and precision stabilized at their greatest value around approximately 20 plots, but study designs with 10 plots performed well (Fig. 14B-E). When study designs had four revisits, accuracy and precision were high for sample sizes as low as five plots (Fig. 14B-E).

Table 12: Simulated data inputs of external biotic homogenization analysis.

Input	UPn/UPs	FDc/FDn	MHc/MHs
baseline mean	0.8417	0.7532	0.8529
plot pair variance (g_p)	0.003	0.0053	0.0048
residual variance (e_{pc})	0.01	0.01	0.01
completed plots	9/33	6/34	23/38

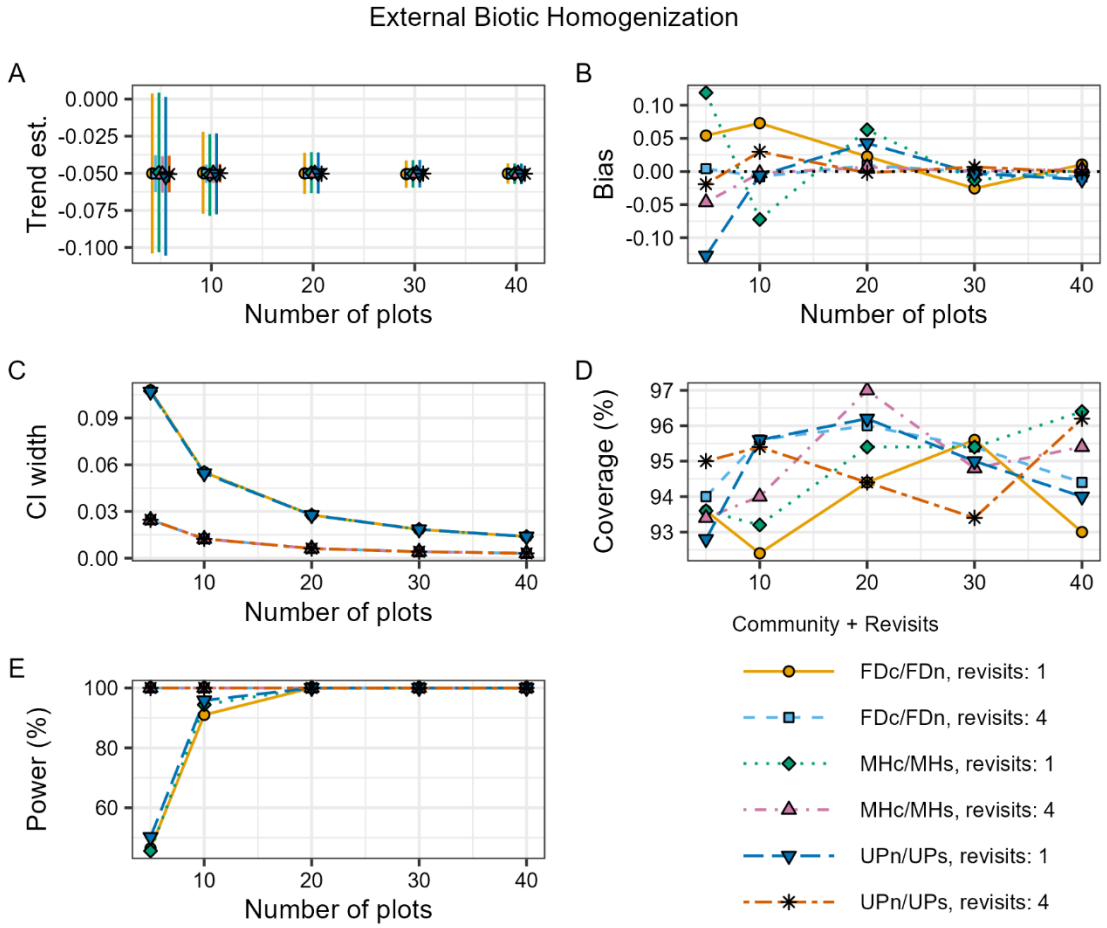


Figure 14: Power analysis of external biotic homogenization. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Forest impacts

Deer browsing

The existing EMN forest data suggests that for every percentage increase in browse pressure, the log-odds of *Acer* occurrence increased by 5% (Table 13). Simulated datasets with effect sizes of this magnitude had estimates within 1% for samples sizes of at least 10 plots and power of at least 87% for sample sizes of 200 plots (Fig. 15B, E). Coverage was relatively high across study designs (Fig. 15D). When the effect size was increased to +/- 10%, power of at least 87% could be reached with approximately 50 plots (Fig. 15E). However, effect sizes of -/+ 1% had power less than 25% across all study designs.

Table 13: Simulated data inputs of deer browse analysis.

Input	Value
baseline mean	0.51
residual variance	12.74
predictor mean ± sd	0.37 ± 0.48
observed effect	5%
observed P-value	0.002
completed plots	170
non-random trend	0

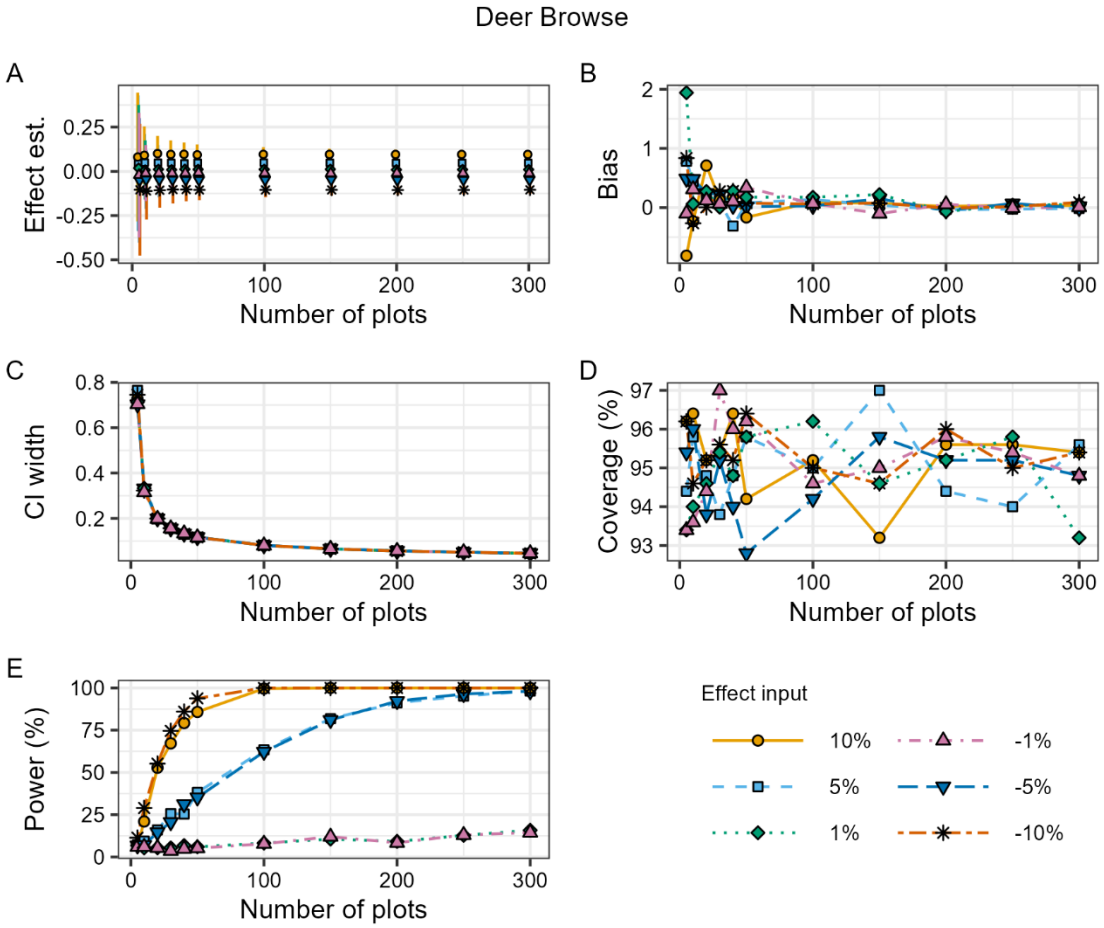


Figure 15: Power analysis of deer browse pressure effects on Acer spp. occurrence. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

When we assumed 200 plots had been sampled, varying the proportion of random plots slightly affected accuracy and cover (Fig. 16A). Bias was greatest (but still within 1%) when only 10% of the plots were random (Fig. 16B). Coverage exceeded 90% and power exceeded 72% when 150 (or 3/4ths) of the 200 plots were chosen randomly (Fig. 16D-E).

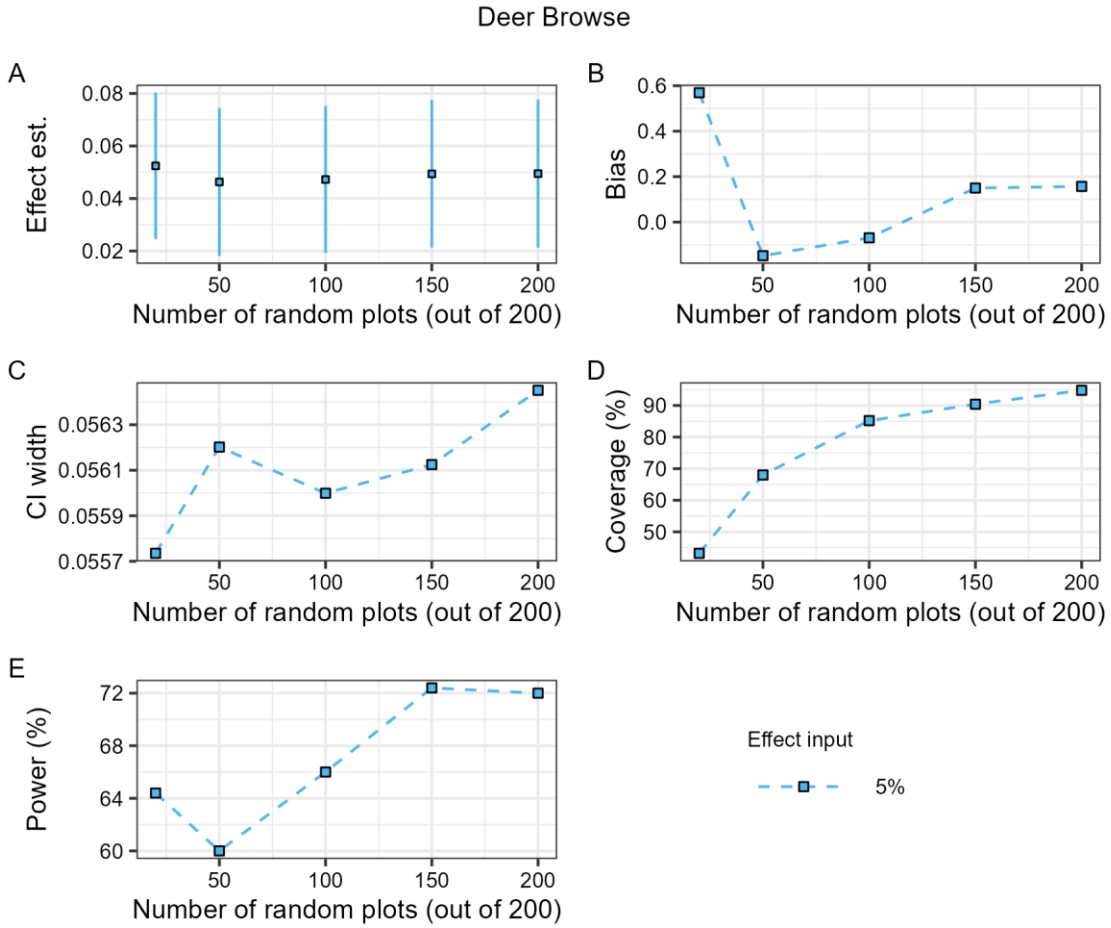


Figure 16: Power analysis of deer browse pressure effects on Acer spp. occurrence with incremented inclusion of non-random plots (out of 200). (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Earthworms

Data collected from EMN forest plots suggest there was no statistically significant effect of earthworm disturbance on graminoid cover (Table 14). The ordinal nature of the earthworm disturbance variable prevented us from assessing study designs with five plots because models could not be fit to simulated datasets with too few unique values (Fig. 17A). While coverage was relatively high across study designs (Fig. 17D), estimates deviated substantially from input effect sizes and were not consistently within 1% until study designs had at least 100 plots (Fig. 17B). CI widths were also relatively large until study designs had at least 100 plots (Fig. 17C). Power was consistently low and did not exceed 60% across all study designs and effect sizes (Fig. 17E).

Table 14: Simulated data inputs of earthworm disturbance analysis.

Input	Value
baseline mean	0.07
residual variance	1.75
predictor mean ± sd	0.02 ± 0.14
observed effect	-5%
observed P-value	0.432
completed plots	147

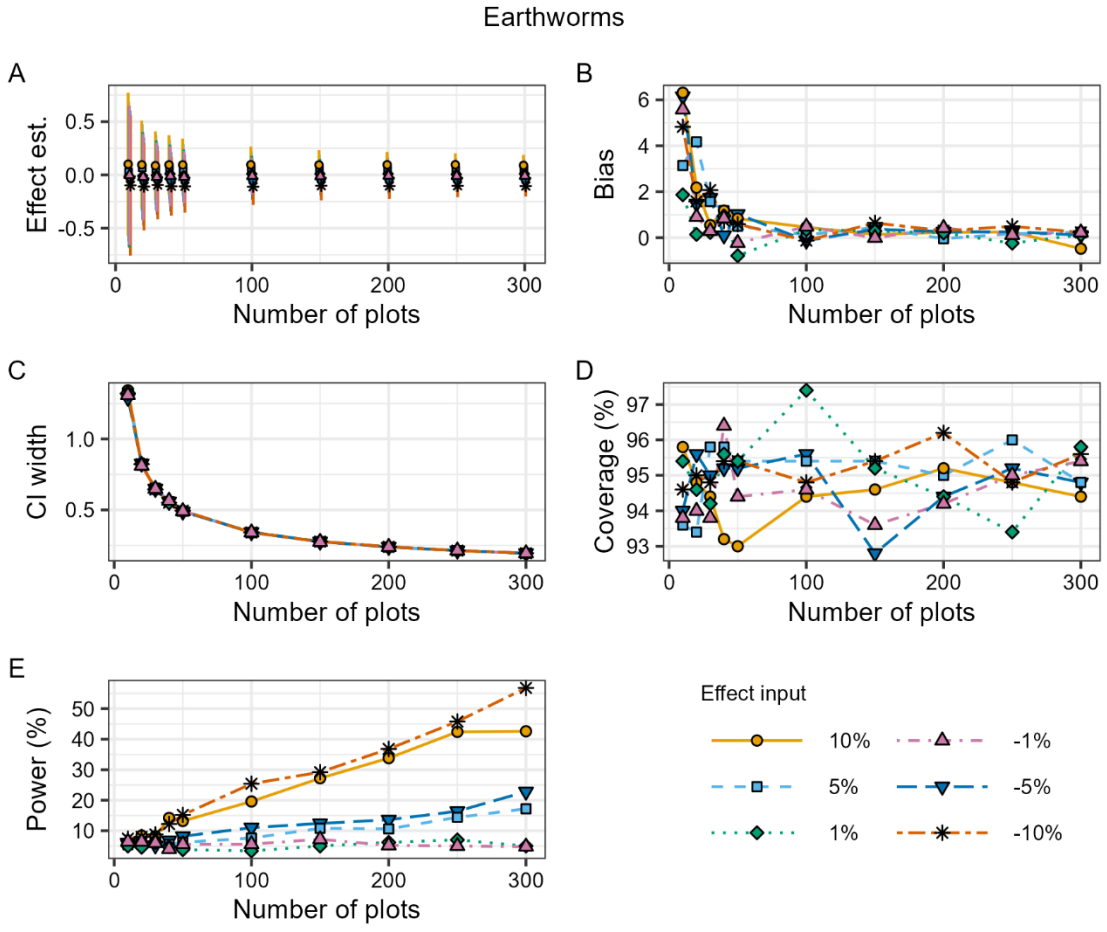


Figure 17: Power analysis of earthworm disturbance effects on graminoid cover. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Timber harvesting

We do not yet have estimates for the effects of timber harvesting on basal area in EMN forest plots, but the average basal area of *Acer* species could be estimated from 127 plots (Table 15). Accuracy and precision of timber harvesting models increased with sample sizes (Fig. 18). Effect size estimates were consistently within 1% of true values for sample sizes of at least 30 plots and coverage was consistently high across study designs (Fig. 18B, D). CI width was less than 0.25 for at least 100 plots and power was at least 87% for 150 plots (Fig. 18C, E).

Table 15: Simulated data inputs of timber harvest analysis.

Input	Value
baseline mean	171.79
residual variance	1.34
predictor mean	10
completed plots	127

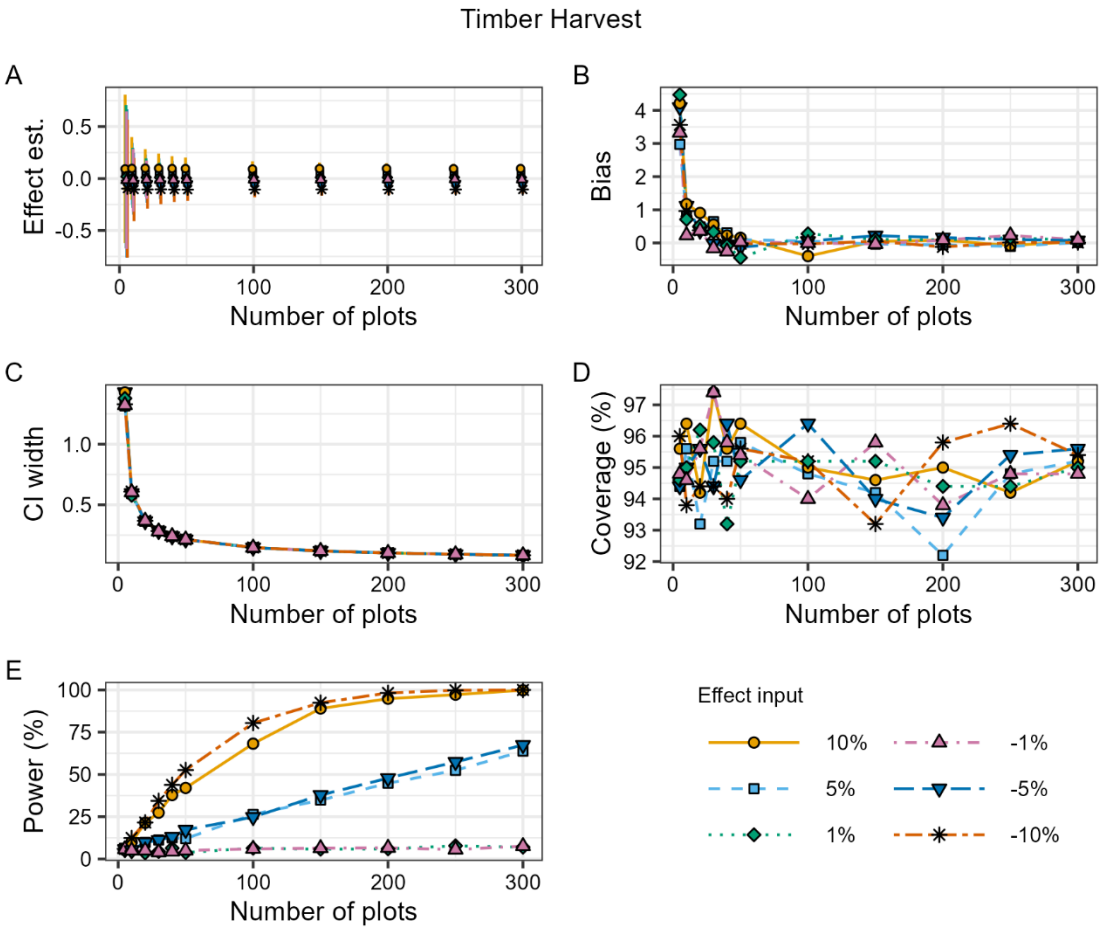


Figure 18: Power analysis of timber harvest effects on *Acer* spp. average basal area. (A) Trend estimates on the log-odds scale, averaged across simulations, with 95% confidence intervals (CI). Points and CI are nudged horizontally to reduce overlap. (B) Average difference between trend estimate and trend input on the percentage scale. (C) Average width of 95% CIs on the log-odds scale. (D) Percentage of models with 95% CIs that include trend input. (E) Percentage of models with $P < 0.05$.

Appendices

Appendix A: Sample Selection

EMN's Method

To make inferences about a population, the samples should be a "miniature population". GRTS survey design can produce samples that are representative of the population because they are random and spatially balanced across the population (Olsen et al. 2012). The GRTS draw for the EMN study design included 20,000 points with the state of Minnesota as the sampling frame (Rowe 2022). During plot evaluation, landowners were called to request access for the project. If a point was excluded, the reason was recorded. However, sites were not evaluated in order of the GRTS draw.

EMN plots were chosen with some discretion using the GRTS draw of 20,000 points. A large proportion of the GRTS points are in non-target areas (land or water that is not native plant community). It has been more efficient to focus on areas where native plant communities are known to exist or where field work can be conducted efficiently and choose GRTS points in those areas rather than to evaluate the points in the order supplied by the algorithm. Additionally, GRTS points may land near, but not in a native plant community. Therefore, some plot locations have been moved to capture native plant communities.

It is also desirable to intentionally include some plot locations in EMN. Because EMN is funded by the state legislature and aims to represent the entire state, it is important to have plots in counties throughout the state. However, some counties have relatively small areas of native plant communities and so plot locations need to be chosen deliberately. In addition, some types of plant communities are rare or tend to be found on cliff faces, reducing the probability that they'll be represented in the GRTS sample and requiring deliberate selection. Finally, collaborators may be interested in collecting vegetation data at their site and request an EMN plot be placed there.

Analysis considerations

Plots selected following the GRTS survey design can be analyzed with design-based approaches, which require no assumptions about the population being sampled (Lookingbill et al. 2012) and can take point exclusion into account through design weights (Dumelle et al. 2022). Design weights are used for scaling inference from the sample to the population (Stevens and Olsen 2004). Further, the spatial location of the points can be used to estimate local variance, which tends to be smaller than non-spatial variance estimates (Stevens and Olsen 2003). Smaller variance can lead to smaller sample size requirements to detect a given trend with a given amount of power. We did not use local variance in the analyses because it was unclear which plots were selected strictly following the GRTS design. A hybrid model-based/design-based approach could eventually be used to estimate local variance (Brus and de Gruijter 2012, Dumelle et al. 2022). In addition, design can be incorporated into linear regression models to account for samples that are mixes of randomly and non-randomly chosen points (Irvine et al. 2018, Starcevich et al. 2018), as demonstrated with the unequal weight model used to analyze multiple indicators in this report.

Deviations from the order of the GRTS algorithm are non-random because the person selecting points is making some decision about where and how to select. For points selected out of order, the GRTS survey design can be restored by evaluating all the skipped points and including target points in the EMN sample. Plot locations that were selected within a given buffer area of the GRTS point may be considered as following the design. At some distance from the points though, the location is being selected non-randomly. How do we determine this distance? One approach is to determine the minimum distance between any two GRTS points in the subset of consecutive points that are evaluated for EMN. Divide this distance by two to get a buffer radius that can be considered as "belonging" to a GRTS point. A plot location within this buffer radius may then be considered as part of the GRTS survey design, assuming it was chosen in order.

While plots selected out of order or due to stakeholder interests no longer adhere to the GRTS survey design, they can still be analyzed with the regression methods described in this report. Separating out randomly and non-randomly selected plots in this context and assigning them design weights may not be possible due to the many decisions used to select plots. Therefore, results using unweighted regression methods can be presented with appropriate context about how plots were chosen. Insights from researchers about how estimates for trend values and effect sizes may deviate from those that would result if GRTS points were sampled in order would help the audience interpret how plot selection methods affected the results. Specific plots can be omitted from analyses if they were chosen through distinct methods and researchers suspect they deviate strongly from the target population.

Interpretation considerations

Consistent deviations from the GRTS algorithm are evidence that the sampling frame does not represent the population of interest or was not properly stratified (e.g., increasing inclusion probability of points in the Prairie Parkland Province, where native plant communities are rare, relative to the Laurentian Mixed Forest Province). At the time of study design, the population could not be accurately represented by GIS resources to develop a sampling frame, so EMN researchers have been essentially modifying the sampling frame post-hoc through their decisions about which GRTS points are in the target population. When data or results from EMN are presented in documents or talks, the population of interest should be described precisely. In addition, the links between the population description and GRTS points decisions should be outlined.

The specific plot location is determined on-the-ground when the sampling crew arrives at the GRTS point location. The criteria used to select the specific plot locations are described by Rowe (2022), which can be referenced when data or results are presented in documents or talks. Deviations from these criteria should be noted.

Appendix B: Analyzing Community Composition

R packages such as Hmsc (Tikhonov et al. 2020) allow users to fit joint species distribution models, which can be non-spatial, and estimate effect sizes for each taxon. They assume correlations among species to account for biotic interactions and concurrent responses to environmental drivers that are not otherwise included in the model. Finally, they have the flexibility to evaluate drivers of community change, and incorporate trait and phylogenetic information as a means to evaluate the relationships between drivers and species' responses (Tikhonov et al. 2020). A joint species distribution model was not used for power analysis because the Hmsc

package uses Markov Chain Monte Carlo sampling to estimate model parameters, which is prohibitively slow when paired with the many iterations needed for power analysis. There are quicker methods for fitting joint species distribution models, such as the gllvm package (Niku et al. 2019). Future community composition analyses using EMN data should explore these and other available tools.

Appendix C: Generic indicators with alternative baseline averages

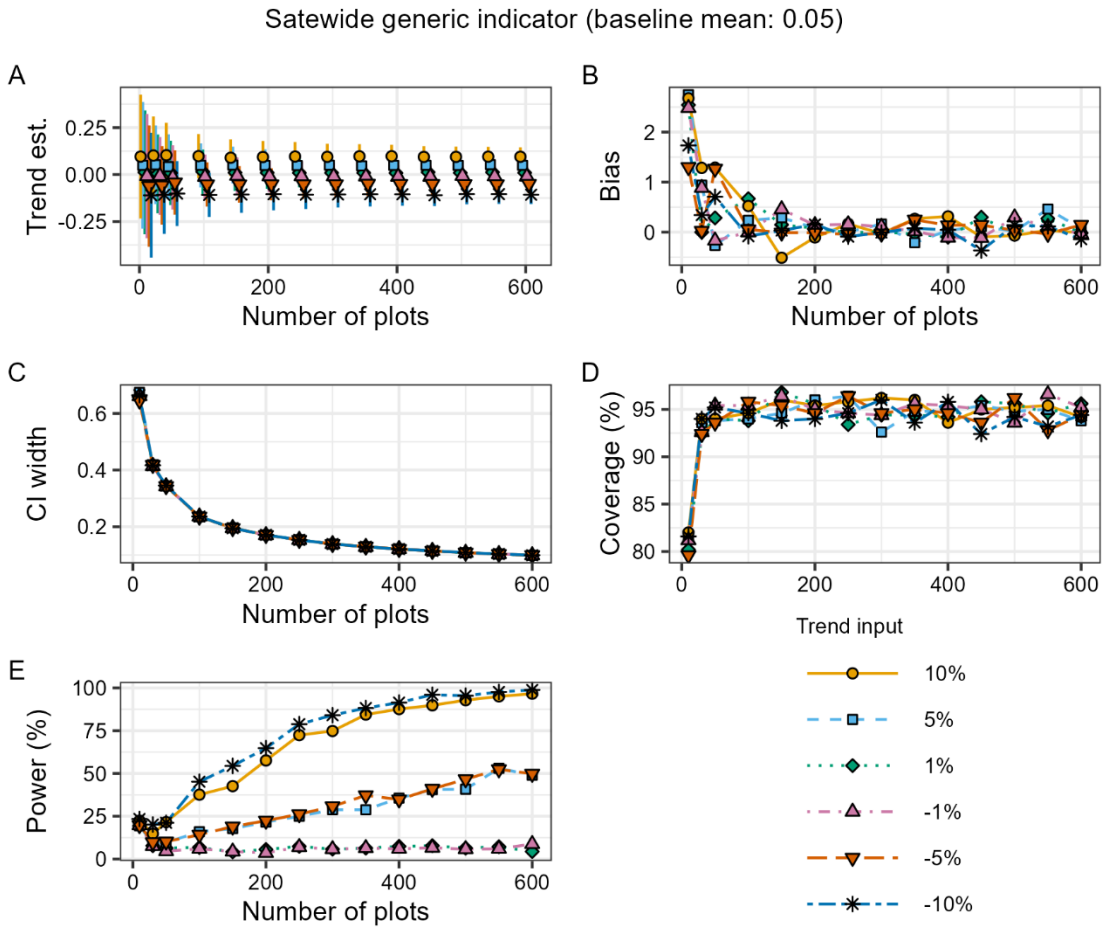


Figure 19: Power analysis of statewide generic indicator with lower baseline average.

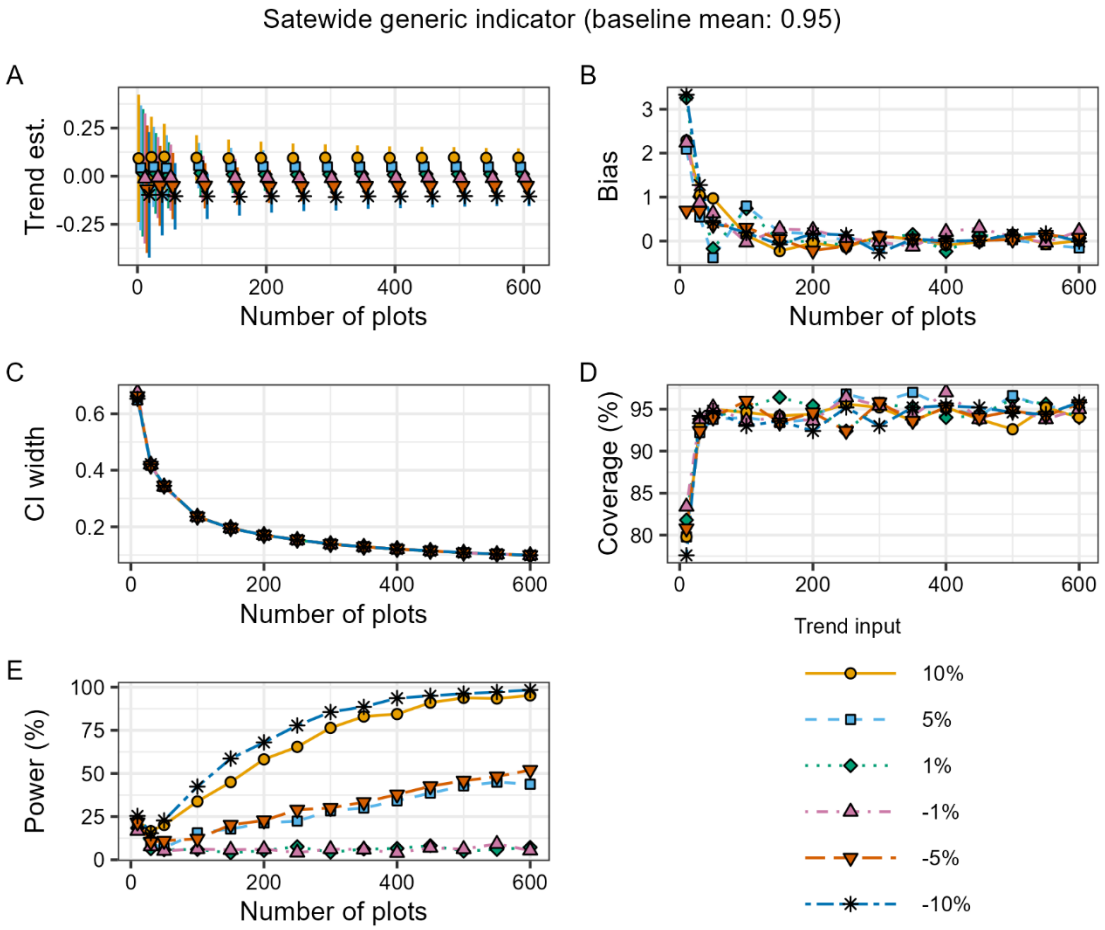


Figure 20: Power analysis of statewide generic indicator with higher baseline average.

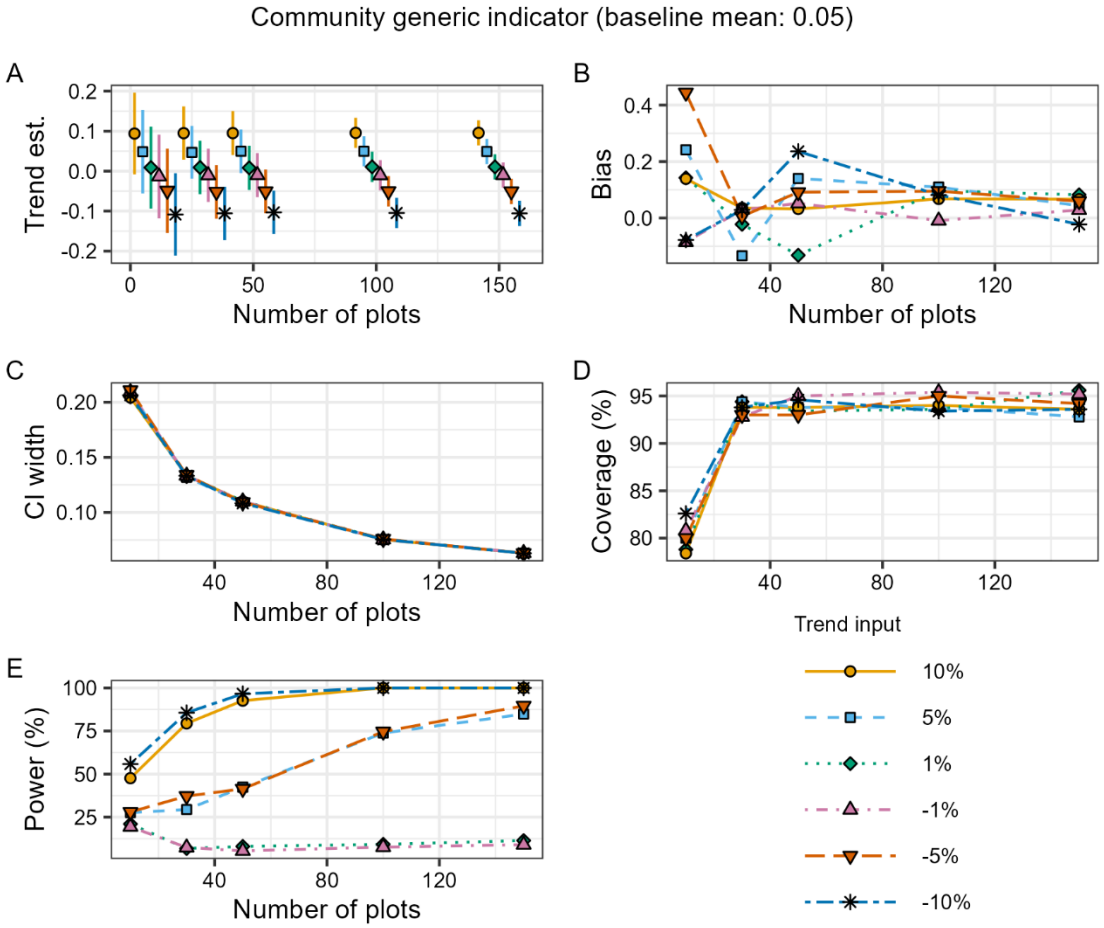


Figure 21: Power analysis of community-scale generic indicator with lower baseline average.

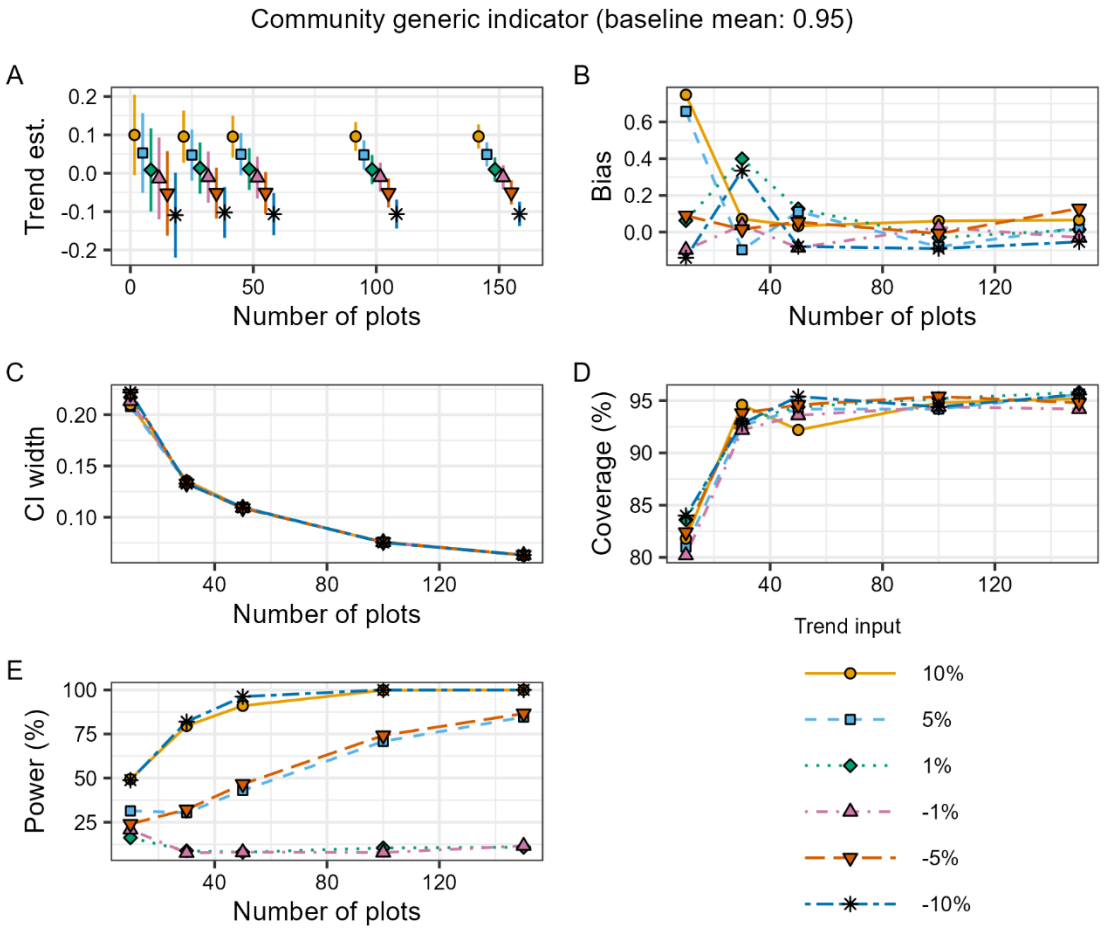


Figure 22: Power analysis of community-scale generic indicator with higher baseline average.

References

Bolker, B. M. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.

Bouchard, M., and D. Pothier. 2011. Long-term influence of fire and harvesting on boreal forest age structure and forest composition in eastern Québec. Forest Ecology and Management 261:811–820.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal 9:378–400.

Brus, D. J., and J. J. de Gruijter. 2012. A hybrid design-based and model-based sampling approach to estimate the temporal trend of spatial means. Geoderma 173–174:241–248.

Dumelle, M., T. M. Kincaid, A. R. Olsen, and M. H. Weber. 2022. spsurvey: Spatial Sampling Design and Analysis.

Fisichelli, N. A., L. E. Frelich, P. B. Reich, and N. Eisenhauer. 2013. Linking direct and indirect pathways mediating earthworms, deer, and understory composition in Great Lakes forests. Biological Invasions 15:1057–1066.

Frerker, K., G. Sonnier, and D. M. Waller. 2013. Browsing rates and ratios provide reliable indices of ungulate impacts on forest plant communities. Forest Ecology and Management 291:55–64.

Irvine, K. M., E. C. Dinger, and D. Sarr. 2011. A power analysis for multivariate tests of temporal trend in species composition. Ecology 92:1879–1886.

Irvine, K. M., and T. J. Rodhouse. 2010. Power analysis for trend in ordinal cover classes: implications for long-term vegetation monitoring: Power analysis for trend in ordinal cover classes. Journal of Vegetation Science 21:1152–1161.

Irvine, K. M., T. J. Rodhouse, W. J. Wright, and A. R. Olsen. 2018. Occupancy modeling species–environment relationships with non-ignorable survey designs. Ecological Applications 28:1616–1625.

Johnson, S. E., E. L. Mudrak, E. A. Beever, S. Sanders, and D. M. Waller. 2008. Comparing power among three sampling methods for monitoring forest vegetation. Canadian Journal of Forest Research 38:143–156.

Johnson, S. E., E. L. Mudrak, and D. M. Waller. 2006. A comparison of sampling methodologies for long-term forest vegetation monitoring in the Great Lakes Network National Parks. Great Lakes Inventory and Monitoring Network Report, National Park Service.

Keddy, P. A., and D. C. Laughlin. 2021. A Framework for Community Ecology: Species Pools, Filters and Traits. First edition. Cambridge University Press.

Lookingbill, T. R., J. P. Schmit, and S. L. Carter. 2012. GRTS and graphs: Monitoring natural resources in urban landscapes. Pages 361–380 *in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.

Loss, S. R., R. M. Hueffmeier, C. M. Hale, G. E. Host, G. Sjerven, and L. E. Frelich. 2013. Earthworm Invasions in Northern Hardwood Forests: a Rapid Assessment Method. Natural Areas Journal 33:21–30.

Miller, K. M., and B. R. Mitchell. 2014. A new tool for power analysis of fixed plot data: Using simulations and mixed effects models to evaluate forest metrics. Ecosphere 5:art110.

MN DNR. 2003. Field Guide to the Native Plant Communities of Minnesota: The Laurentian Mixed Forest Province. Ecological Land Classification Program, Minnesota County Biological Survey, and Natural Heritage and Nongame Research Program, St. Paul, MN.

MN DNR. 2005a. Field Guide to the Native Plant Communities of Minnesota: The Prairie Parkland and Tallgrass Aspen Parklands Provinces. Ecological Land Classification Program, Minnesota County Biological Survey, and Natural Heritage and Nongame Research Program, St. Paul, MN.

MN DNR. 2005b. Field Guide to the Native Plant Communities of Minnesota: The Eastern Broadleaf Forest Province. Ecological Land Classification Program, Minnesota County Biological Survey, and Natural Heritage and Nongame Research Program, St. Paul, MN.

MN DNR. 2022. MNTaxa: The State of Minnesota Vascular Plant Checklist. https://www.dnr.state.mn.us/eco/mbs/plant-lists.html.

Nahorniak, M., D. P. Larsen, C. Volk, and C. E. Jordan. 2015. Using inverse probability bootstrap sampling to eliminate sample induced bias in model based analysis of unequal probability samples. PLOS ONE 10:e0131765.

Niku, J., F. K. C. Hui, S. Taskinen, and D. I. Warton. 2019. gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in R. Methods in Ecology and Evolution 10:2173–2182.

Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. D. Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, D. McGlinn, M.-H. Ouellette, E. R. Cunha, T. Smith, A. Stier, C. J. F. T. Braak, and J. Weedon. 2022. vegan: Community Ecology Package.

Olden, J. D., and T. P. Rooney. 2006. On defining and quantifying biotic homogenization: Defining biotic homogenization. Global Ecology and Biogeography 15:113–120.

Olsen, A. R., T. M. Kincaid, and Q. Payton. 2012. Spatially balanced designs for natural resources. Pages 126–150 *in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.

Perles, S. J., D. R. Manning, K. K. Callahan, and M. R. Marshall. 2014a. Forest Health Monitoring in the Eastern Rivers and Mountains Network: 2009-2012 Summary Report. Natural Resource Report NPS/ERMN/NRR, National Park Service U.S. Department of the Interior, University Park, PA.

Perles, S. J., T. Wagner, B. J. Irwin, D. R. Manning, K. K. Callahan, and M. R. Marshall. 2014b. Evaluation of a regional monitoring program's statistical power to detect temporal trends in forest health indicators. Environmental Management 54:641–655.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. Version 4.4.1. R Foundation for Statistical Computing, Vienna, Austria.

Rowe, E. 2022. Minnesota Ecological Monitoring Network (EMN): Background and Standard Operating Procedures. Minnesota Department of Natural Resources.

Starcevich, L. A., G. DiDonato, T. McDonald, and J. Mitchell. 2016. A GRTS User's Manual for the SDrawNPS Package: A graphical user interface for Generalized Random Tessellation Stratified (GRTS) sampling and estimation. Natural Resource Report NPS/PWRO/NRR, Western EcoSystems Technology, Inc.

Starcevich, L. A. H., T. McDonald, A. Chung-MacCoubrey, A. Heard, J. Nesmith, and T. Philippi. 2018. Trend estimation for complex survey designs of water chemistry indicators from Sierra Nevada Lakes. Environmental Monitoring and Assessment 190:596.

Stevens, D. L., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14:593–610.

Stevens, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262–278.

Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehikoinen, M. M. J. Jonge, J. Oksanen, and O. Ovaskainen. 2020. Joint species distribution modelling with the R -package H MSC. Methods in Ecology and Evolution 11:442–447.

Urquhart, N. S. 2012. The role of monitoring design in detecting trend in long-term ecological monitoring studies. Pages 361–380 *in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.

Urquhart, N. S., and T. M. Kincaid. 1999. Designs for detecting trend from repeated surveys of ecological resources. Journal of Agricultural, Biological, and Environmental Statistics 4:404.

van der Veen, B., F. K. C. Hui, K. A. Hovstad, and R. B. O'Hara. 2023. Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling. Methods in Ecology and Evolution 14:683–695.

Zobel, M. 2016. The species pool concept as a framework for studying patterns of plant diversity. Journal of Vegetation Science 27:8–18.