Environ Monit Assess (2012) 184:5601–5611
DOI 10.1007/s10661-011-2365-3

Detecting change in advance tree regeneration using forest inventory data: the implications of type II error

James A. Westfall · William H. McWilliams

Received: 25 January 2011 / Accepted: 30 August 2011 / Published online: 20 September 2011

Abstract Achieving adequate and desirable forest regeneration is necessary for maintaining native tree species and forest composition. Advance tree seedling and sapling regeneration is the basis of the next stand and serves as an indicator of future composition. The Pennsylvania Regeneration Study was implemented statewide to monitor regeneration on a subset of Forest Inventory and Analysis plots measured by the U.S. Forest Service. As management techniques are implemented to improve advance regeneration, assessments of the change in the forest resource are needed. When the primary focus is on detecting change, hypothesis tests should have small type II (β) error rates. However, most analyses are based on minimizing type I (α) error rates and type II error rates can be quite large. When type II error rates are high, actual improvements in regeneration can remain undetected and the methods that brought these improvements may be deemed ineffective. The difficulty in detecting significant change in advance regeneration when small type I error rates are given priority is illustrated. For statewide assessments, power (1-β) to detect changes in proportion of area having adequate advance regeneration is relatively weak (≤0.5) when the change is smaller than 0.05. For evaluations conducted at smaller spatial scales, such as wildlife management units, the reduced sample size results in only marginal power even when relatively large changes (≥0.20) in area proportion occur. For fixed sample sizes, analysts can consider accepting larger type I error rates to increase the probability of detecting change (smaller type II error rates) when it occurs, such that management methods that positively affect regeneration can be identified.

Keywords Type II error · Tree seedling · Species composition · Forest inventory

J. A. Westfall (✉) · W. H. McWilliams
U.S. Forest Service, Northern Research Station,
11 Campus Boulevard, Suite 200,
Newtown Square, PA 19073, USA
e-mail: jameswestfall@fs.fed.us

Introduction

Preservation and maintenance of forested land often is a high priority for environmental conservation and is contingent upon diversity in native tree species composition and tree size distributions (Liebhold et al. 1995). In most situations, native tree species composition is maintained through naturally occurring regeneration such that the same species are present in both the overstory and as regeneration. To help determine whether forests are replacing themselves, silviculturists often examine advance tree regeneration (Marquis et al. 1992). Mechanisms capable of altering the number and composition of regenerating species include removal of overstory trees due to weather extremes (Fajvan et al. 2006) or harvest/mortality (Kays et al. 1988), competition (Fei et al. 2010; Harmer 2001), deer browsing (Powers and Nagel 2009; Horsley et al. 2003), soil chemistry/pollution (Drohan and Sharpe 1997), and fire (Nowacki and Abrams 2008). When sufficient regeneration is lacking in a given landscape, it is often difficult to attribute failure to a specific causal agent.

Advance regeneration often is monitored through standard forest inventory methods, i.e., a network of sample plots (Johnstone et al. 2004; McWilliams et al. 1995). As such, the planning process entails computing sample size for a desired level of precision. Sample size estimates are often based on estimates of population variability, selection of an acceptable type I error rate, and desired confidence interval width (Schreuder et al. 2004). This approach emphasizes the precision of estimates for current regeneration status. Often overlooked is the ability to detect change when change has indeed occurred—also known as the statistical power to detect change.

As a reminder to readers, types I and II error rates are often defined within the framework of statistical hypothesis testing. In the context of this paper, there is a null hypothesis representing the status quo (or no change) and a research hypothesis that represents some level of change in a population parameter. A type I error occurs when the actual change in the population does not cross the threshold value specified in the research hypothesis, but the research hypothesis is accepted as true due to the sample statistics. Similarly, a type II error is committed when the actual change in the population crosses the threshold value specified in the research hypothesis, but the null hypothesis is accepted as true (Di Stefano 2001). In practice, commission of either type of error remains unidentified as the true population values are unknown. As such, low types I and II error rates are sought to minimize the probability of drawing erroneous conclusions.

The conundrum faced by analysts is that type I (α) and type II (β) errors are inversely related, such that for a given sample size, β increases if α decreases and vice-versa (Kleinbaum et al. 2008). An appropriate balance between α and β error rates requires evaluation of the negative consequences involved with each type of error. Techniques for making these assessments include a cost ratio approach where the costs of each type of error are determined and the ratio of the two error rates (e.g., α/β) is made to correspond with the ratio of respective costs (Peterman 1990). Field et al. (2004) evaluate a cost function that varies with α and β error rates, where the α/β relationship is determined via statistical models. Theoretically, the minimization of the cost function provides the optimal error rates. Other methodologies for determining appropriate values for α and β are described by Murphy and Myors (2004). Generally, type II errors are considered more egregious in environmental monitoring studies because failing to detect (and more importantly respond to) changing conditions is often more deleterious than taking action to correct a perceived change that did not actually occur (type I error; Fairweather 1991; Mapstone 1995). In the context of regeneration, identification of changes in regenerative capabilities of forests is paramount to conservation of indigenous tree species and forest types.

The importance of considering type II error rates for environmental studies has been well-documented; however applications specific to forest vegetation monitoring are infrequent. Evans and Viengkham (2001) considered statistical power analysis to investigate potential survey designs for monitoring Lao rattan. This study seems to have motivated Archaux and Bergès (2008) to use power analysis to optimize a sampling design for detection of a specified level of change in species richness for plant communities. Thompson et al. (2011) evaluated power to detect change in vegetation cover when developing a sampling design for national parks in Alaska. The strength of three different vegetation sampling methods to detect change in species richness, plant abundance, and overstory basal area and composition was studied by Johnson et al. (2008). Bechtold et al. (2009) assessed ability of a forest health monitoring network to detect prescribed levels of change in various tree crown attributes. The relative recency of these papers suggests that although the need for analyses of type II error rates has been recognized for some time; practical applications to forest vegetation monitoring are only beginning to be undertaken.

In this study, data from an existing regeneration monitoring program were used to examine type II error rates, i.e., incorrectly accept the null hypothesis of no change when change has actually occurred. The objectives of this study are: (1) introduce and provide estimation methodology for an area-based metric of regeneration success (proportion of forestland area having adequate regeneration), (2) assess population and subpopulation variability and investigate ability to detect various amounts of change at current and alternative sample sizes, (3) provide information to assist future development of advance tree regeneration surveys.

Methods

Data

In cooperation with the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service, the Pennsylvania Regeneration Study (PRS) was initiated in 2001 by the Pennsylvania Department of Conservation and Natural Resources, Bureau of Forestry (McWilliams et al. 2003a, b) to capture critical measurements not included in core FIA protocols. The regeneration study was incorporated into the existing FIA sample and plot design, where the sampling intensity is approximately one plot for every 6,000 acres of area (Reams et al. 2005). Each sample plot consists of four 24-ft radius subplots, and within each subplot is a 6.8-ft radius microplot (Bechtold and Scott 2005). All trees with a diameter breast height (dbh) of 5.0 in. and larger are measured on the subplot, while information on saplings (1.0 in.≤dbh< 5.0 in.) and seedlings are collected on the microplot. For the PRS, 1,549 plots (one plot per ~18,000 acres) were systematically selected from the existing FIA sample to maintain the statewide spatial distribution. This sample size was chosen based on field work considerations (e.g., summer only) and assumed state-level analytical output. On this subset of plots, additional information was collected on forested lands (USDA Forest Service 2007). In addition to the regular suite of FIA variables, measurements included a location-level evaluation of deer impact (as determined from browse intensity), seedling height class and number, and percent cover of competing vegetation such as ferns, shrubs, and grasses (McWilliams et al. 2002). At the time of this analysis, 807 plots had received a second measurement (5-year interval).

Although the survey was implemented at the state level, analyses at smaller spatial scales usually are more useful. In 2003, a Deer Management Plan (Pennsylvania Game Commission 2003) was instituted to better manage for quality deer habitat across Wildlife Management Units (WMUs). As such, analyses are most useful when performed for subpopulations defined by WMU boundaries. There are 22 WMUs statewide with areas ranging from approximately 0.5–2.7 million acres (average area is roughly 1.3 million acres).

Analysis

For both the initial and subsequent measurement, selected tree species regeneration within a range of sizes occurring on microplots with 45–70% overstory stocking density were the domain of interest (for further explanation, see definitions of $\delta_{ij\mathrm{FA}}$ and $\delta_{ij\mathrm{F}}$ below). The overstory stocking constraint defines a density where favorable light conditions exist for healthy understory development of native canopy species (McWilliams et al. 1995). There were 244 species of interest—those considered capable of forming a high canopy (a detailed list is available upon request). For seedlings, the relative importance of each stem was gauged by assigning a weight based on its height. Seedlings were assigned to the following height classes: <1 ft, 1.0–2.9 ft, 3.0–4.9 ft, and ≥5.0 ft; with weights for each height class being 1, 2, 20, and 50, respectively, i.e., larger seedlings are more important in assessing adequate regeneration (McWilliams et al. 1995). The weights were chosen using published regeneration guidelines (Sander et al. 1976; Marquis and Bjorkbom 1982) to work in accordance with the regeneration thresholds used in the SILVAH decision support program (Marquis et al. 1992). The number of weighted stems needs to be assessed in the context of herbivory pressure. As such, a microplot was considered adequately stocked with regeneration if the sum of the weighted seedlings exceeded the index established for the level of deer impact (Table 1). If the sum of weighted seedlings did not exceed the threshold, but there existed a sapling having dbh 1.0–4.9 in., the microplot was considered to have sufficient regeneration. Large-seeded species (e.g., oak) were coded as "established" or "competitive" when root-collar diameter attained 0.75 in. or larger. Research has shown that root-collar diameter is the best indicator of success, particularly under high deer impact or fire scenarios (Brose 2008).

The cluster-plot design used by FIA (Bechtold and Scott 2005) requires that results from each microplot i

Table 1 Weighted seedlings thresholds for sufficient advance regeneration by level of deer impact

Deer impact	Weighted seedlings threshold
Very low	15
Low	30
Medium	50
High	100
Very high	200

be summarized to a single observation for each plot j. This summary is dependent upon the variable of interest. If the variable of interest was forestland area having 40–75% overstory stocking, the summary would be the proportion of total microplot area that contains forestland within the prescribed stocking range. Similarly, for area having sufficient regeneration, the summary would be the proportion of total microplot area that is forested, 40–75% stocked, and has adequate regeneration. For this paper, these are respectively defined by:

$$y_j = \sum_{i=1}^{4} \delta_{ijFA}/4$$

$$x_j = \sum_{i=1}^{4} \delta_{ijF}/4$$

where

y_j Proportion of plot j that is forested, 40–75% stocked, adequate regeneration

x_j Proportion of plot j that is forested and 40–75% stocked

$\delta_{ij\mathrm{FA}}$ $\begin{cases} 1 & \text{if microplot } i \text{ on plot } j \text{ is forested and} \\ & 40-75\% \text{ stocked, adequate ATSSR} \\ 0 & \text{otherwise} \end{cases}$

$\delta_{ij\mathrm{F}}$ $\begin{cases} 1 & \text{if microplot } i \text{ on plot } j \text{ is forested and} \\ & 40-75\% \text{ stocked} \\ 0 & \text{otherwise} \end{cases}$

The ratio-of-means estimator of the proportion of 40–75% stocked forestland that has sufficient regeneration, $\widehat{P}_R$, is equivalent to the mean proportion of plot area that is 40–75% stocked forestland having adequate regeneration divided by the mean proportion of plot area that is 40–75% stocked forestland.

$$\widehat{P}_R = \frac{\frac{\sum y_j}{n}}{\frac{\sum x_j}{n}} = \frac{\overline{y}}{\overline{x}}$$

The correlation between y_j and x_j due to both measurements occurring on the same sample plot are accounted for in the estimated variance (Cochran 1977),

$$V(\widehat{P}_R) = \frac{1}{n\overline{x}^2}\left(s_y^2 + \widehat{P}_R^2 s_x^2 - 2\widehat{P}_R s_{yx}\right)$$

where

n Sample size

s_y^2 Sample variance of the y_j

s_x^2 Sample variance of the x_j

s_{yx} Sample covariance between the y_j and x_j

For computation of change in proportion of forested area having adequate regeneration, measurements at two points in time on the same sample plots are used. With the estimation procedures outlined above, the proportion of forestland having adequate regeneration at each time can be calculated ($\widehat{P}_{R1}$ and $\widehat{P}_{R2}$, respectively), and the change over the time period is,

$$\widehat{P}_{R\Delta} = \widehat{P}_{R2} - \widehat{P}_{R1}$$

with variance estimator that accounts for the correlation due to using the same sample plots at each time (Cochran 1977),

$$V(\widehat{P}_{R\Delta}) = V(\widehat{P}_{R2}) + V(\widehat{P}_{R1}) - 2\mathrm{Cov}(\widehat{P}_{R2}, \widehat{P}_{R1})$$

where

$$\mathrm{Cov}(\widehat{P}_{R2}, \widehat{P}_{R1}) = \frac{1}{n(n-1)\overline{x}_2\overline{x}_1} \sum_{j=1}^{n} (y_{j2}y_{j1} - \widehat{P}_{R2}y_{j1}x_{j2} - \widehat{P}_{R1}y_{j2}x_{j1} + \widehat{P}_{R2}\widehat{P}_{R1}x_{j2}x_{j1})$$

The standard error of the estimate is given by: $\mathrm{SE}(\widehat{P}_{R\Delta}) = \sqrt{V(\widehat{P}_{R\Delta})}$

Having computed the estimate of change and the associated error, the next step is to make inferences regarding change occurring in the population. The primary test is whether there was a statistically significant change in proportion of area of forestland having adequate regeneration between the two time periods. There are two types of erroneous conclusions to avoid: (1) type I error determining the change in the proportion was statistically different from zero when actually it was not, and (2) type II error-resolving there was no change in the proportion when a change did occur (Di Stefano 2001). The type I error rate (α) and type II error rate (β) represent the probability of drawing the respective incorrect conclusions. The statistical power of a test (1-β) is the probability that true change in the proportion of forest area having sufficient regeneration will be detected. The probability depends on the magnitude of the change $\widehat{P}_{R\Delta}$, α, n, the null mean, and the standard deviation (s) of the sample data (Foster 2001). The null mean is the value to which $\widehat{P}_{R\Delta}$ is being compared, most commonly zero (no change). The standard deviation was calculated as $s = \mathrm{SE}\left(\widehat{P}_{R\Delta}\right)\sqrt{n}$. Analyses were conducted at two spatial scales: a statewide assessment to evaluate overall conditions, and an assessment for individual WMUs because this is the scale at which the results are most useful to wildlife managers. Table 2 provides a summary of pertinent statistics by WMU and for the entire state.

The SAS POWER procedure (SAS Institute Inc. 2008) was used to analyze the observed data. Additional power analyses were conducted to inform what-if scenarios and provide readers with a broader view of the relationships between the statistical power and the various inter-related parameters.

Results

Statewide

With the data provided in Table 2, the power to detect change in proportion of forestland with adequate regeneration was evaluated for the state as well as for individual WMUs. At the state level, the power to detect the change in proportion of forestland having adequate regeneration of −0.002 (assuming the traditional α=0.05) as statistically different from zero was 0.05, i.e., a type II error would be committed in 95% of the samples. These results should also be interpreted in the context of practical vs. statistical significance. Given that −0.002 is very close to zero, one may consider −0.002≈0 for all practical purposes. If one subscribes to this approximate equality, there is no actual change in the proportion of area having adequate regeneration and type II errors are not of concern. It should be noted that it would be rare to have a forest resource survey that indicates change is exactly zero, thus type II errors do need to be considered for most applications.

It also is instructive to evaluate the power in relation to other factors. The power to detect significant change from zero is decreased to 0.01 when α=0.01 and it is increased to 0.10 when α= 0.10. The assumption that −0.002≈0 helps illustrate a particular point of interest-the minimum power attainable is equal to α and this occurs at the null value being tested against (often zero). Another factor is the difference between the sample estimate (−0.002) and the null value to be tested against. For example, the deer population may have increased and thus managers expected the proportion of area having adequate regeneration to change by −0.05. As such, it may be desirable to test whether the estimate from the sample is different from the expected change. For α= 0.05, the results show that the power to detect a significant difference under this scenario is 0.30. This means that if a number of independent samples were conducted, a type II error would be committed in 70% of the samples, i.e., the hypothesis $\widehat{P}_{R\Delta} = -0.05$ could not be rejected. Figure 1 depicts the relationships between statistical power at various α levels and null values for the statewide evaluation.

A more general question of interest is how much change in forestland area with adequate regeneration must occur to conclude the change is significant. This analysis uses zero for the null mean and Figure 1 depicts this relationship in the case when the observed change approaches zero (i.e., −0.002≈0). With the statewide sample size and standard deviation (s), the power to detect significant changes in proportion of forestland having adequate regeneration $\left(\widehat{P}_{R\Delta}\right)$ differs substantially with α. When α=0.01, the power to detect a change of −0.03 is 0.04. In contrast, the power to detect this same change improves to 0.34 when α=0.20. Generally, the power is relatively weak (≤0.5) for detecting $\widehat{P}_{R\Delta} \neq 0$ when the actual change is smaller than −0.05. Obtaining statistical power in the 0.7–0.8 range requires an actual change in area proportion of 0.06–0.07.

Table 2 Summary of data for WMU and statewide for both initial and subsequent measurements with estimates of change ($\widehat{P}_{R\Delta}$) and standard deviation (s)

WMU	n	Time 1			Time 2			$\widehat{P}_{R\Delta}$	s
		Forestland proportion	Forestland proportion (40–75%)	$\widehat{P}_{R1}$	Forestland proportion	Forestland proportion (40–75%)	$\widehat{P}_{R2}$		
1A	32	0.490	0.304	0.378	0.491	0.266	0.419	0.041	0.778
1B	36	0.431	0.222	0.453	0.438	0.188	0.398	−0.055	0.921
2A	30	0.562	0.394	0.480	0.565	0.333	0.413	−0.068	0.982
2B	22	0.250	0.182	0.609	0.250	0.148	0.712	0.102	0.879
2C	52	0.673	0.404	0.557	0.670	0.420	0.577	0.020	0.953
2D	47	0.592	0.335	0.520	0.603	0.325	0.443	−0.077	0.849
2E	23	0.548	0.374	0.638	0.568	0.250	0.489	−0.149	1.232
2F	44	0.892	0.455	0.294	0.881	0.421	0.358	0.064	0.578
2G	76	0.833	0.476	0.443	0.836	0.467	0.364	−0.079	0.629
3A	22	0.636	0.227	0.550	0.636	0.443	0.603	0.053	1.474
3B	42	0.750	0.476	0.472	0.755	0.564	0.527	0.056	0.780
3C	36	0.729	0.313	0.433	0.750	0.438	0.552	0.118	1.035
3D	35	0.817	0.571	0.588	0.809	0.486	0.540	−0.047	0.758
4A	30	0.617	0.325	0.635	0.617	0.392	0.590	−0.044	0.899
4B	32	0.562	0.267	0.674	0.547	0.281	0.576	−0.098	1.366
4C	34	0.559	0.235	0.500	0.544	0.228	0.484	−0.016	1.491
4D	50	0.627	0.250	0.405	0.632	0.273	0.421	0.016	1.116
4E	33	0.313	0.194	0.556	0.316	0.238	0.484	−0.072	0.950
5A	24	0.179	0.135	0.519	0.167	0.083	0.750	0.231	1.320
5B	57	0.211	0.118	0.407	0.211	0.053	0.250	−0.157	2.544
5C	36	0.274	0.189	0.221	0.247	0.132	0.461	0.240	1.518
5D	14	0.000	0.000	–	0.000	0.000	–	–	–
Statewide	807	0.564	0.314	0.482	0.564	0.311	0.480	−0.002	0.960

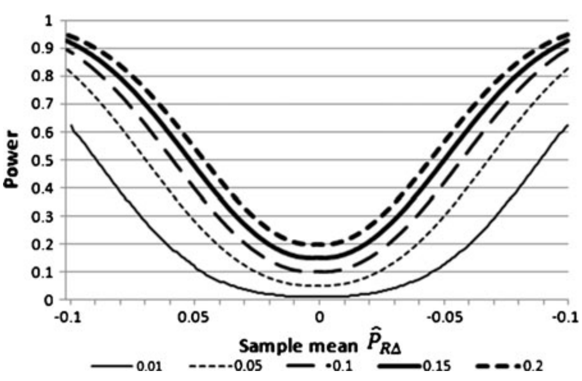

Fig. 1 Power to detect change in proportion of area having adequate regeneration ($\widehat{P}_{R\Delta}$) for selected type I error rates (α= 0.01, 0.05, 0.1, 0.15, and 0.2) using 807 remeasured plots in Pennsylvania

Besides selection of the α value, the primary factor involved in improving power is sample size. In this study, 807 remeasured plots were analyzed, and in the next few years the sample will increase to approximately 1,500 plots with completion of the second cycle of measurements. Knowing this, we can estimate the statistical power for change detection when all plots have been remeasured. Figure 2 shows the power to detect $\widehat{P}_{R\Delta} \neq 0$ when α=0.05 for a range of sample sizes and $\widehat{P}_{R\Delta}$. The power to detect a change of −0.05 increases from 0.30 to 0.52 when the sample size increases from 807 to 1,500. If the number of plots involved in the study increased to 2,000, the power to detect a change of −0.05 would be 0.64 and would be 0.81 for 3000 plots. With 3,000 plots, a type I error rate of 0.05 would yield a type II

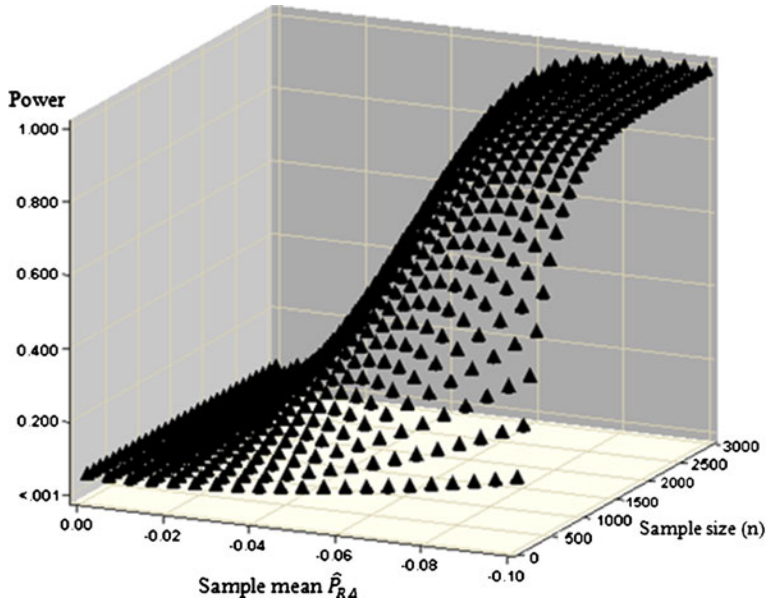

Fig. 2 Power to detect change in proportion of area having adequate regeneration ($\hat{P}_{R\Delta}$) at various sample sizes (n) using the statewide standard deviation (s) estimate=0.96 and type I error rate (α)=0.05. Note that a symmetric pattern would be found for positive sample means

error rate of about 0.20 when $\hat{P}_{R\Delta} = -0.05$. From an operational standpoint, doubling the sample size would require relatively large additional expense, which may not be feasible.

Wildlife management unit

The results among WMUs exhibited high variability (Table 2), primarily due to specific WMU characteristics and as an artifact of the decreased sample sizes attributable to the smaller spatial extent. The $\hat{P}_{R\Delta}$ values were mostly within ±0.10, however the range of values was from −0.157 to 0.240. The more extreme values occurred in WMU having relatively small amounts of forestland (5A, 5B, 5C). Values for s covered a range of approximately 0.5–1.5, with many values near 1.0. An evaluation of power to detect $\hat{P}_{R\Delta} \neq 0$ generally applicable at the WMU spatial scale was conducted assuming s=1.0 and n=40 (Figure 3). For the most commonly selected alpha levels, such as α=0.05 or α=0.10, only marginal power is obtained even when relatively large changes in proportion of forestland having adequate regeneration ($\hat{P}_{R\Delta} \geq 0.20$) are found in the data. Although power is increased at larger α levels (α=0.20), the type II error will still be about 0.3 when a $\hat{P}_{R\Delta} = 0.1$ is obtained from the sample.

Perhaps the most important evaluation to consider at the WMU scale is how the relatively small sample sizes affect power to detect $\hat{P}_{R\Delta} \neq 0$. Assuming s=1.0, the ability to detect significant differences from zero for $\hat{P}_{R\Delta} = -0.05$ and $\hat{P}_{R\Delta} = -0.10$ was evaluated for various α levels (Figure 4). Given that most of the WMU sample sizes are below 50 plots, it is unlikely that a conclusion of $\hat{P}_{R\Delta} \neq 0$ will be drawn for the magnitude of $\hat{P}_{R\Delta}$ found in these data. Figure 4a shows that power to detect $\hat{P}_{R\Delta} = -0.05$ as being statistically different from zero is less than 0.2 for most α values. Only slight improvements are realized for detecting $\hat{P}_{R\Delta} = -0.10$ as being statistically different from zero (Figure 4b), where power will be less than 0.3 in most cases. These are general results applicable at the WMU scale. WMUs having

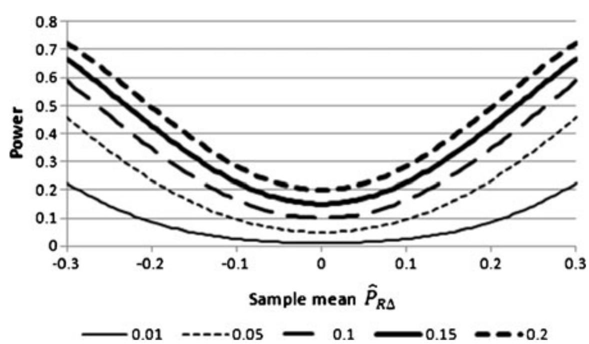

Fig. 3 Power to detect change in proportion of area having adequate regeneration ($\hat{P}_{R\Delta}$) for selected type I error rates (α=0.01, 0.05, 0.1, 0.15, and 0.2) when the standard deviation (s)=1.0 and the sample size (n)=40

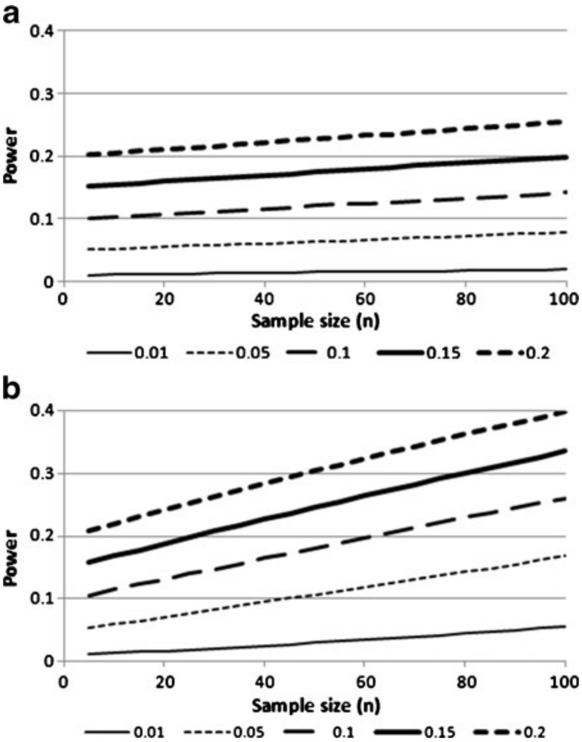

Fig. 4 Power to detect change in proportion of area having adequate regeneration ($\widehat{P}_{R\Delta}$) at various sample sizes (n) for selected type I error rates (α=0.01, 0.05, 0.1, 0.15, and 0.2) and the standard deviation (s)=1.0 when **a** $\widehat{P}_{R\Delta}=-0.05$, and **b** $\widehat{P}_{R\Delta}=-0.10$

relatively small s and large n will have higher probabilities of detecting change. For example, in WMU 2G where $\widehat{P}_{R\Delta}\approx-0.08$, s=0.629, and n=76 plots, the power to avoid a type II error is nearly 0.44 for α=0.20.

Discussion

Assessing current regeneration status is important for understanding the present state of the forest ecosystem and the long-term ability to sustain native forests. In response, landscape-level management activities designed to manipulate regeneration are implemented, e.g., control of browsing populations, herbicides, or prescribed fire (Brose 2008; Loftis 1985). To assess the effectiveness of these activities, a test against a null mean of zero (no change) is appropriate. In the context of type I errors, an erroneous conclusion would result in the continuation of a management strategy that had no effect. In the case of a type II error, a strategy that truly had an impact would be deemed ineffective and likely discontinued. Thus, continued implementation of strategies that genuinely influence regeneration success requires avoidance of type II errors.

Because of the management implications for high type II error rates, more attention should be given to statistical power in the study planning process. Results shown here can provide guidance on appropriate sample sizes for detecting various levels of change in proportion of area where regeneration is satisfactory. Assessments for "average" populations (s≈1.0), when $\widehat{P}_{R\Delta}=0.05$ is considered a level of change that should be detected, indicates approximately 2300 plots would be needed to conclude that $\widehat{P}_{R\Delta}\neq 0$ if acceptable error rates were α=0.1 and β= 0.2 (power=0.8). Obtaining these same error rates for other populations requires evaluation for the given amount of variability (s). In this study, the smallest standard deviation (s=0.578) would require 800 plots, whereas the largest standard deviation encountered (s=1.491) and would require nearly 5,500 plots (note the areal extent of the population of interest does not influence the sample size calculation). These two extremes illustrate the wide range of sampling requirements that may be encountered for different populations. When designing the study, these statistics need to be considered in the context of resources available to conduct the study. If the desired sample sizes are unattainable, the power for reduced sample sizes should be calculated and a reassessment of the study plan undertaken. For instance, if population boundaries are expected to remain unchanged over time, one may consider different sampling intensities among populations; with more samples being devoted to areas with high variability. Another option would be to reconsider the requisite error rates—perhaps the initial specification was a best-case scenario, while a smaller amount of statistical power may be acceptable. Ultimately, a determination needs to be made as to the wisdom of conducting the study given the power attainable.

In the selection of a sample size, a critical piece of information that will often be missing is the standard deviation (s). It was noted during the course of analyses that a relationship existed between the standard deviation (s) and the proportion of the area that contained forestland. The nonlinear correlation as described by the regression of forestland proportion on s via a power function is shown in Figure 5 ($R^2 \approx 0.33$). Generally, the value of s increases as the proportion of forestland area decreases. One reason may be that populations with smaller forested areas are more likely to be fragmented and regeneration on these fragments is more variable than for larger, more contiguous forest areas. The proportion forestland relationship with s may be unduly influenced by the point representing $s \approx 2.5$; however, there is no reason to believe it is not a valid observation. Nonetheless, the trend provides general guidance for survey planning purposes. More accurate estimates of standard deviation for new study areas may be obtained from other sources of information or a pilot study.

When sample sizes are already fixed, analysts should attempt to strike an appropriate balance between error rates and the objectives of the study. In many cases, selection of the α value is the primary focus, while β is nearly always ignored. Furthermore, the focus on α has resulted in the adoption of α=0.05 for most research analyses (Di Stefano 2001). Given that β increases as α decreases, setting a small value for α often results in limited ability to avoid type II errors. In the context of regeneration, it could be postulated that prevention of type II errors is of equal (or perhaps greater) importance than avoidance of type I errors. In the case of a type I error, ineffective management strategies would be continued despite no actual change in area having adequate regeneration. In contrast, a type II error would likely result in abandonment of management practices that were genuinely improving resource conditions. Analysts should weigh the resulting negative social, economic, and environmental impacts associated with each of these errors to determine whether the primary emphasis should be on α or β when other factors are already predetermined.

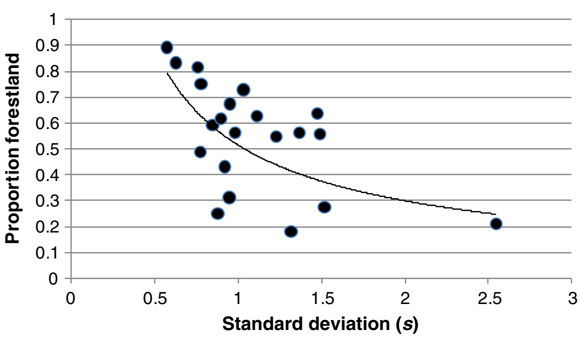

Fig. 5 Relationship between proportion of forestland and standard deviation (s) for WMUs in Pennsylvania

It may also be useful to consider alternative metrics that may impart similar interpretation or knowledge, but have different statistical properties than $\widehat{P}_{R\Delta}$. To develop estimates of $\widehat{P}_{R\Delta}$, classification thresholds are applied to each microplot—this results in information loss compared to the continuous variable(s) on which the classification is based. The information loss translates into a reduction of statistical power, which makes detecting differences more difficult (McCaffrey and Elliott 2008). It may be worthwhile to investigate other metrics, such as a direct analysis of weighted stems, to evaluate if any improvement in probabilities of detecting change may be obtained. While the use of other metrics could possibly be advantageous for change detection, meaningful interpretations of these metrics are also needed; for instance, how does one evaluate estimates of change in weighted stems without considering the associated herbivory pressure.

Conclusion

The statewide analyses showed that when change in proportion of forestland having sufficient advance regeneration is less than 0.05, it is likely to go undetected as being statistically significant. The problem is exacerbated when analyses are undertaken for the WMUs, where sample sizes are substantially reduced. For these smaller populations, even substantial shifts ($\widehat{P}_{R\Delta} > 0.20$) are unlikely to be identified as significant change. These outcomes illustrate the importance of balancing sample size with the desired spatial scale of analyses. In this particular case, analysis at the WMU level was implemented long after the study was designed. However, it is a useful reminder that sufficient sample sizes are needed at the desired scale of analytical resolution.

As forest managers employ various management techniques designed to improve regeneration of desirable tree species, continuing assessments of the efficacy of these activities are prudent. Under an adaptive management framework, activities that do not produce intended effects should be compared to alternative methods. However, when type II error rates are high, it is likely that actual improvements in regeneration will go undetected and the methods that lead to these improvements may be dismissed. To enhance the ability to identify methods that have a positive impact on regeneration, analysts should consider accepting larger type I error rates than are typically adopted. This will increase the probability of detecting change (smaller type II error rates) and foster continued implementation of management methods that contribute to regeneration success. As such, analysts should give more attention to the type II error rate during the planning phase, where approximate type I and type II error rates for proposed sample sizes can be fully explored prior to study initiation. For studies already implemented, if alternative metrics can be developed that impart essentially the same information, comparisons should be made to identify those that achieve the smallest error rates.

References

Archaux, F., & Bergès, L. (2008). Optimising vegetation monitoring. A case study in a French lowland forest. *Environmental Monitoring and Assessment, 141*, 19–25.

Bechtold, W. A., & Scott, C. T. (2005). The Forest Inventory and Analysis plot design. In: W. A. Bechtold & P. L. Patterson, (Eds.), *The enhanced forest inventory and analysis program—national sampling design and estimation procedures* (pp. 27–42). General Technical Report SRS-80. Asheville, NC: U.S. Dept. of Agriculture, Southern Research Station.

Bechtold, W. A., Randolph, K. C., & Zarnoch, S. J. (2009). The power of FIA phase 3 crown-indicator variables to detect change. In: W. McWilliams, G. Moisen, R. Czaplewski (Comps.), *Forest Inventory and Analysis (FIA) symposium 2008*. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Dept. of Agriculture, Rocky Mountain Research Station.

Brose, P. H. (2008). Root development of acorn-origin oak seedlings in shelterwood stands on the Appalachian Plateau of northern Pennsylvania: 4-year results. *Forest Ecology and Management, 255*, 3374–3381.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.

Di Stefano, J. (2001). Power analysis and sustainable forest management. *Forest Ecology and Management, 154*, 141–153.

Drohan, J. R., & Sharpe, W. E. (1997). Long-term changes in forest soil acidity in Pennsylvania, USA. *Water, Air, and Soil Pollution, 95*, 299–311.

Evans, T. D., & Viengkham, O. V. (2001). Inventory time-cost and statistical power: A case study of Lao rattan. *Forest Ecology and Management, 150*, 313–322.

Fairweather, P. G. (1991). Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research, 42*, 555–567.

Fajvan, M. A., Plotkin, A. B., & Foster, D. R. (2006). Modeling tree regeneration height growth after an experimental hurricane. *Canadian Journal of Forest Research, 36*, 2003–2014.

Fei, S., Gould, P., Kaeser, M., & Steiner, K. (2010). Distribution and dynamics of the invasive native hay-scented fern. *Weed Science, 58*, 408–412.

Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R., & Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters, 7*, 669–675.

Foster, J. R. (2001). Statistical power in forest monitoring. *Forest Ecology and Management, 151*, 211–222.

Harmer, R. (2001). The effect of plant competition and simulated summer browsing by deer on tree regeneration. *Journal of Applied Ecology, 38*, 1094–1103.

Horsley, S. B., Stout, S. L., & deCalesta, D. S. (2003). White-tailed deer impact on the vegetation dynamics of a northern hardwood forest. *Ecological Applications, 13*, 98–118.

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

Johnstone, J. F., Chapin, F. S., III, Foote, J., Kemmett, S., Price, K., & Viereck, L. (2004). Decadal observations of tree regeneration following fire in boreal forests. *Canadian Journal of Forest Research, 34*, 267–273.

Kays, J. S., Smith, D. W., Zedaker, S. M., & Kreh, R. E. (1988). Factors affecting natural regeneration of Piedmont hardwoods. *Southern Journal of Applied Forestry, 12*, 98–102.

Kleinbaum, D., Kupper, L., Nizam, A., & Muller, K. (2008). *Applied regression analysis and other multivariable methods* (4th ed.). Belmont: Duxbury.

Liebhold, A. M., Macdonald, W. L., Bergdahl, D., & Mastro, V. C. (1995). Invasion by exotic forest pests: A threat to forest ecosystems. *Forest Science Monograph, 30*, 1–49.

Loftis, D. L. (1985). Preharvest herbicide treatment improves regeneration in southern Appalachian hardwoods. *Southern Journal of Applied Forestry, 9*, 177–180.

Mapstone, B. D. (1995). Scalable decision rules for environmental impact studies: Effect size, type i, and type II errors. *Ecological Applications, 5*, 401–410.

Marquis, D. A., & Bjorkbom, J. C. (1982). *Guidelines for evaluating regeneration before and after clearcutting Allegheny hardwoods*. Research Note NE-307. Broomall, PA: U.S. Dept. of Agriculture, Northeastern Research Station.

Marquis, D. A., Ernst, R. L., & Stout, S. L. (1992). *Prescribing silvicultural treatments in hardwood stands of the Alleghenies (revised)*. General Technical Report NE-96. Broomall, PA: U.S. Dept. of Agriculture, Northeastern Research Station.

McCaffrey, D. F., & Elliott, M. N. (2008). Power of tests for a dichotomous independent variable measured with error. *Health Services Research, 43*, 1085–1101.

McWilliams, W. H., Stout, S. L., Bowersox, T. W., & McCormick, L. H. (1995). Adequacy of advance tree-seedling regeneration in Pennsylvania's forests. *Northern Journal of Applied Forestry, 12*, 187–191.

McWilliams, W. H., King, S. L., & Scott, C. T. (2002). Assessing regeneration adequacy in Pennsylvania's forests: A pilot study. In: G. A. Reams, R. E. McRoberts, P. C. Van Deusen, (Eds.), *Proceedings of the second annual Forest Inventory and Analysis symposium* (pp. 119–122). General Technical Report SRS-47. Asheville, NC: U.S. Dept. of Agriculture, Southern Research Station.

McWilliams, W. H., Bowersox, T. W., Brose, P. H., Devlin, D. A., Finley, J. C., Horsley, S., et al. (2003a). Indicators of regenerative capacity for eastern hardwood forests. In *Proceedings of Society of American Foresters 2003 National Convention* (pp. 136–141). Bethesda, MD: Society of American Foresters.

McWilliams, W. H., Bowersox, T. W., Brose, P. H., Devlin, D. A., Finley, J. C., Horsley, S., et al. (2003b). Indicators of regenerative capacity for native broadleaf forests: A case study of ungulate overpopulation in the Eastern United States. In *Proceedings IUFRO World Forestry Congress XII* (p. 342). http://www.fao.org/DOCREP/ARTICLE/WFC/XII/0947-B3.HTM.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah: Erlbaum Associates, Inc.

Nowacki, G. J., & Abrams, M. D. (2008). The demise of fire and "mesophication" of forests in the eastern United States. *BioScience, 58*, 123–138.

Pennsylvania Game Commission. (2003). *Population management plan for white-tailed deer in Pennsylvania (2003–2007).* Harrisburg PA: Pennsylvania Game Commission. http://www.portal.state.pa.us/portal/server.pt/document/715654/management__plan6-03_pdf.

Peterman, R. M. (1990). Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences, 47*, 2–15.

Powers, M. D., & Nagel, L. M. (2009). Pennsylvania sedge cover, forest management and deer density influence tree regeneration dynamics in a northern hardwood forest. *Forestry, 82*, 241–254.

Reams, G. A., Smith, W. D., Hansen, M. H., Bechtold, W. A., Roesch, F. A., & Moisen, G. G. (2005). The forest inventory and analysis sampling frame. In W. A. Bechtold & P. L. Patterson, (Eds.), *The enhanced forest inventory and analysis program – national sampling design and estimation procedures* (pp. 11–26). General Technical Report SRS-80. Asheville, NC: U.S. Dept. of Agriculture, Southern Research Station.

Sander, I. L., Johnson, P. S., & Watt, R. F. (1976). *A guide for evaluating the adequacy of oak advance reproduction.* General Technical Report NC-23. St. Paul, MN: U.S. Dept. of Agriculture, North Central Forest Experiment Station.

SAS Institute Inc. (2008). *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Schreuder, H. T., Ernst, R., & Ramirez-Maldonado, H. (2004). *Statistical techniques for sampling and monitoring natural resources*. General Technical Report RMRS-GTR-126. Fort Collins, CO: U.S. Dept. of Agriculture, Rocky Mountain Research Station.

Thompson, W. L., Miller, A. E., Mortenson, D. C., & Woodward, A. (2011). Developing effective sampling designs for monitoring natural resources in Alaskan national parks: An example using simulations and vegetation data. *Biological Conservation, 144*, 1270–1277.

USDA Forest Service. (2007). *Forest inventory and analysis national core field guide, volume 1: Field data collection procedures for phase 2 plots* (Ver. 4.0). Arlington VA: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis. http://fia.fs.fed.us/library/field-guides-methods-proc/.