

SIEN Merced Riverbank Study: Power Analysis

December, 2012

Matt Nahorniak

Statistical Consultant, University of Idaho Dept. of Statistics

nahornim@onit.orst.edu,

(541) 740-5487

Introduction

This study aims to estimate tradeoffs between site level measurement intensity and precision of site level and reach level metric estimation, for metrics included in the SIEN Merced Riverbank Study. Analysis of existing GRTS sample data from a reach of the Merced river suggests that when reach level estimates are of interest, there is very little improvement in precision of estimates when increasing sub-site sampling intensity beyond about 20 transects per bank per site, or beyond about 10 locations per transect. Changes in the total number of sites sampled will always have an effect on the precision of reach level estimates.

Thirteen metrics (see list, below) have been assessed for 24 sites along the Merced River as part of the SIEN Merced Riverbank Study. The 24 sites were selected by a GRTS sampling design, with two stratifications included in the design. Measurements at each site were conducted in 2010.

Within each site sampled, assessment of each metric at each site is made by taking a large number of transects, at both river banks, at each site; and assessing the metric at a large number of points within each transect. At the site level, the metric estimates describe the proportion of each site that falls into the metric category. At each point along each transect, each metric is evaluated as either present (1) or not-present (0). Site level estimates are obtained by averaging all present (0) and not-present (1) values within each site.

The R-package `spsurvey`, developed specifically for the analysis of data obtained from GRTS designs, is then used to make reach level estimates, based on site level estimates, along with reach level estimates of uncertainty (standard errors and confidence intervals).

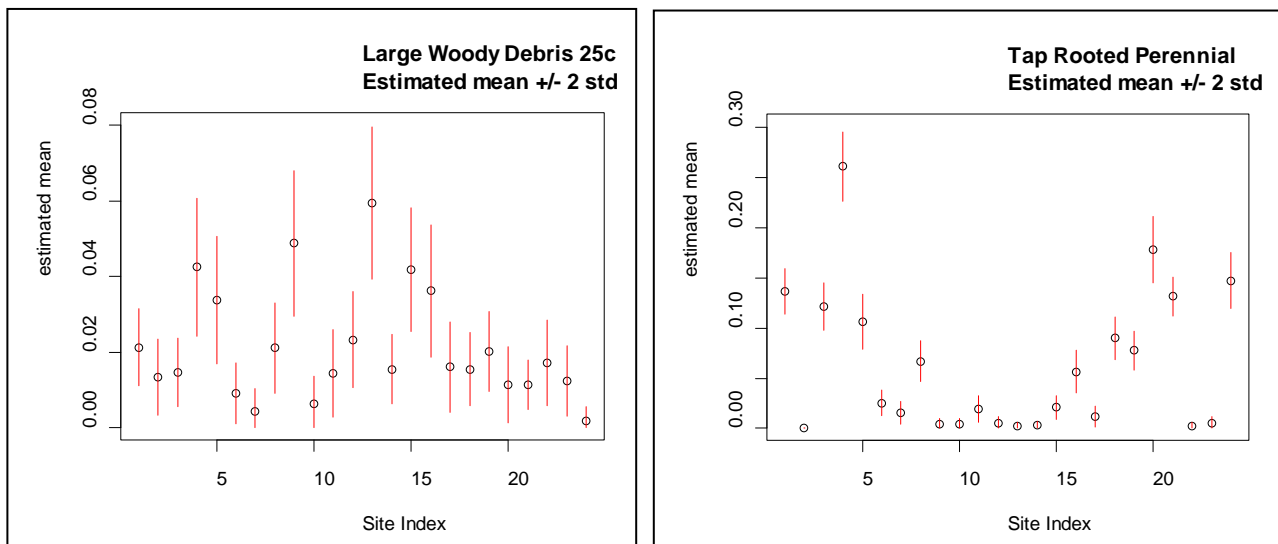
This paper investigates the effect of varying the total number of sites sampled, the number of transects per site, and the number of points assessed per transect, on the precision of both site-level and reach-level estimates for the various metrics assessed. "Reach level" estimates, in this case, are defined as estimates for the weighted average response across all sites included in the sampling frame (i.e. all possible sites in the reach, from which the sample of sites was obtained). For reach level estimates in particular, there is generally a point of sharply diminishing returns as the number of transects and/or points per transect is increased. Resource optimization requires determination of these trends and appropriate resource allocation based on these trends.

The following metrics are included in this analysis:

- Bare Ground
- Litter 25 cm diameter
- Large Woody Debris 25cm diameter
- Exposed Roots
- Non Vascular Plant
- Annual Biennial
- Fibrous Root Rhizomatous Perennial
- Tap Rooted Perennial
- Shrub
- Woody Seedling 0.5m tall
- Live Canopy Evergreen Tree
- Live Canopy Deciduous Tree
- Snag Dead Standing Trees

Background

In general, efficient spatial sampling maximizes sampling units over the spatial level that contains the highest level of variability. For a given metric, assuming site-site variability is high relative to within site variability, a high number of sites sampled are required to achieve high precision for reach level metric estimates. Figures 1 and 2 show examples of site-site variability for two metrics included this analysis: Large Woody Debris and Tap Rooted Perennial. Note the high level of site-site variability.



High site-site variability is common for spatially varying metrics, as points near each other are often more alike than points further apart in space. Hypothetically, if one were to increase the number of

sites sampled (regardless of the number of points measured within each site), the estimated reach level mean would converge on the true reach level mean as the number of sites increased to infinity. However, if a fixed number of sites were sampled, the reach level estimated mean will not converge to the true reach level mean, even if the number of transects and/or points sampled per transect approached infinity. At some point, increasing the number of transects and/or points sampled per transect will result in little or no improvement in precision of reach level estimates.

Plots showing site-site variability for all 13 metrics considered in this analysis are included in Appendix A1.

Within Site Sampling Intensity: Simulation Methodology

Within sites, sampling at different numbers of transects per site, and different numbers of points per transect, was simulated using bootstrapping, where resampling, with replacement, was done using actual 2010 raw, sub-site level data.

For the purposes of this study, data from both banks of the river was used, for all sites that included data for both banks in the original dataset. The effect on precision of sampling only 1 bank versus sampling both banks was not considered, as bank-bank variability is always assumed large relative to transect-transect variability within a bank, or point-point variability within a transect. Sampling only one bank would not make sense from a resource optimization perspective, given that sampling transects from both banks does not entail much, if any, additional effort compared to sampling the same number of transects on only one bank. This appears to be the case for the Merced River, given the easy access to both banks of the river. For the remainder of this analysis and report, the number of transects sampled is always the number of transects *per bank*, unless otherwise noted. Thus the total site level effort, in terms of number of points sampled per site, is:

$$\text{Sampling Intensity: Total Points/Site} = \text{Transects/Bank} * \text{Points/transect} * 2 \text{ Banks/Site}$$

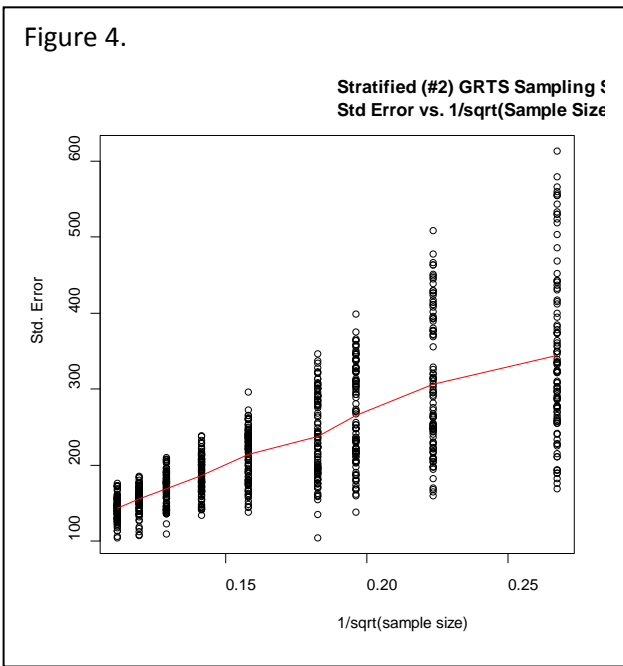
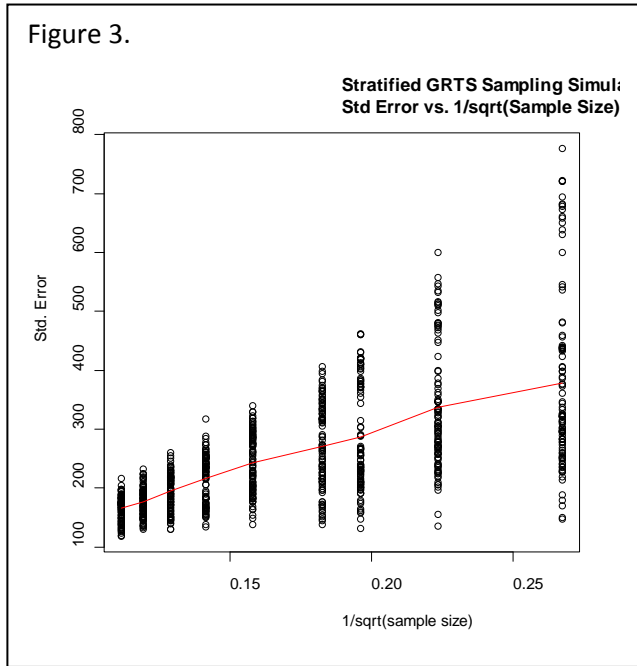
Reach Level Sampling Intensity: Simulation Methodology

At the multi-site level, conventional bootstrap methods are not appropriate for estimating the change in standard errors at various numbers of total sites sampled. This is because a GRTS sample is an ordered list, and the ordering is necessary to preserve the spatially balanced properties. A bootstrap sampling of sites at the watershed level would result in a random sample that lacked the spatial balance of the original sample. Therefore, in lieu of bootstrapping, a different approach is necessary to estimate the effect of changes in total number of sites surveyed on reach level standard errors.

In simple random sampling, well established theory shows that the standard error scales linearly with $1/\sqrt{n}$, where n is the total sample size. A heuristic argument can be made that an appealing

property of a GRTS sample is that, while the standard error isn't straightforward to calculate (as in a simple random sample of independent observations) the standard error should nevertheless retain the linear relationship with $1/\sqrt{n}$. For a given sample size (where sample size is referring to number of sites), assume we have a known or estimated standard error (in this case based on analysis of the existing GRTS design using Spsurvey). The value of that standard error will be a function of the variance and spatial distribution of the metric across sites, and the sampling plan (stratifications, panels, etc.) Regardless of the spatial distribution and GRTS sampling design, adding additional sites to an existing design adds them in a continuously spatially balanced manner, across any known or unknown spatial variations, across the entire sampling frame – provided the additional sites added were sampled using the same GRTS design. This differs from, say, a block sample where n_1 blocks are sampled and n_2 sites are sampled within blocks. In this case, adding changing n_1 or n_2 would not necessarily cause the standard error to scale with $1/\sqrt{n}$, because the increase in sampling intensity might not be linear at all spatial levels.

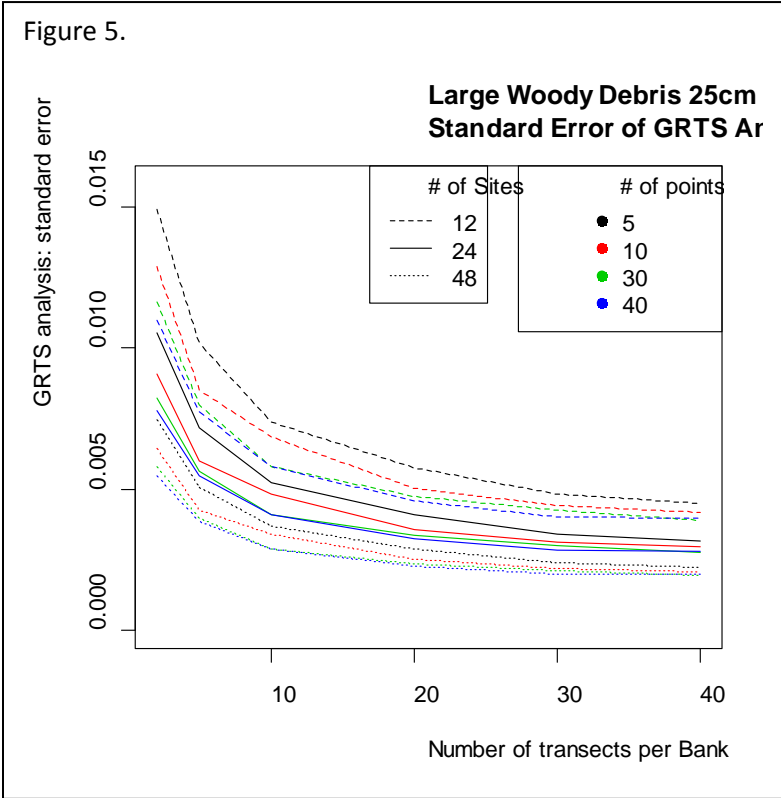
The above heuristic argument was tested empirically using a sampling frame and dataset obtained from the EPA aquatic resources monitoring homepage (<http://www.epa.gov/nheerl/arm/index.htm>). This dataset included spatial locations, and various measurements or other properties, covering all sites within a sampling frame covering the Luckiamute Watershed in the Willamette Valley of Oregon. Using the data frame, and treating various measurement or other properties as site level response measurements, multiple GRTS designs (Equal probability and 2 versions of stratified samples) were created, and repeatedly simulated and analyzed using `cont.analysis` from the `spsurvey` R package. The sampling design and analysis were repeated 1000's of times for each GRTS design, and at various total samples sizes for each GRTS design. The resulting standard errors, for each GRTS design, were plotted against $1/\sqrt{n}$. In each case, over all metrics considered, the plots were consistent with a linear relationship. Two such plots are shown by the Figures 3 and 4:



Figures 3 and 4 support the assertion that standard error from an analysis of GRTS sample data is at least approximately linear with $1/\sqrt{n}$, within a given GRTS design. This assertion will be relied on for the results presented in the next section.

Results

For the Large Woody Debris metric, Figure 5 provides a set of curves where the response (along the vertical axis) is the standard error of the reach level proportion of Large Woody Debris. The x-axis is the number of transects per site per bank; the color of each line shows the number of points per transect observed, and the type of line (solid, dotted, dashed) corresponds to the total number of sites sampled within the reach of the Merced River. Figure 6 shows the same curves for the proportion of Tap Rooted Perennial.

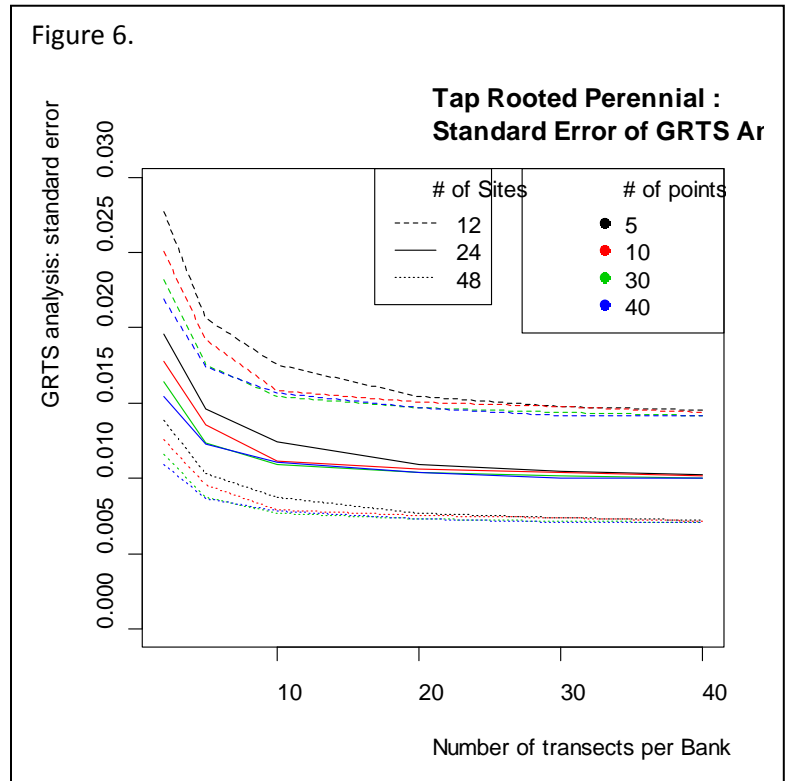


Standard error is a measure of uncertainty in results. It is an inverse quantity to precision. Thus, in Figures 5 and 6, lower standard error indicates higher precision.

As would be expected, increasing the number of sites sampled always increases the precision of the estimates at the reach level.

Increasing the number of transects per site per bank also increases the reach level precision (reduces the reach level standard error), but the improvement in precision drops considerably after roughly 30 transects per bank for Large Woody Debris, and after as few as 10 transects per site for Tap Rooted Perennial.

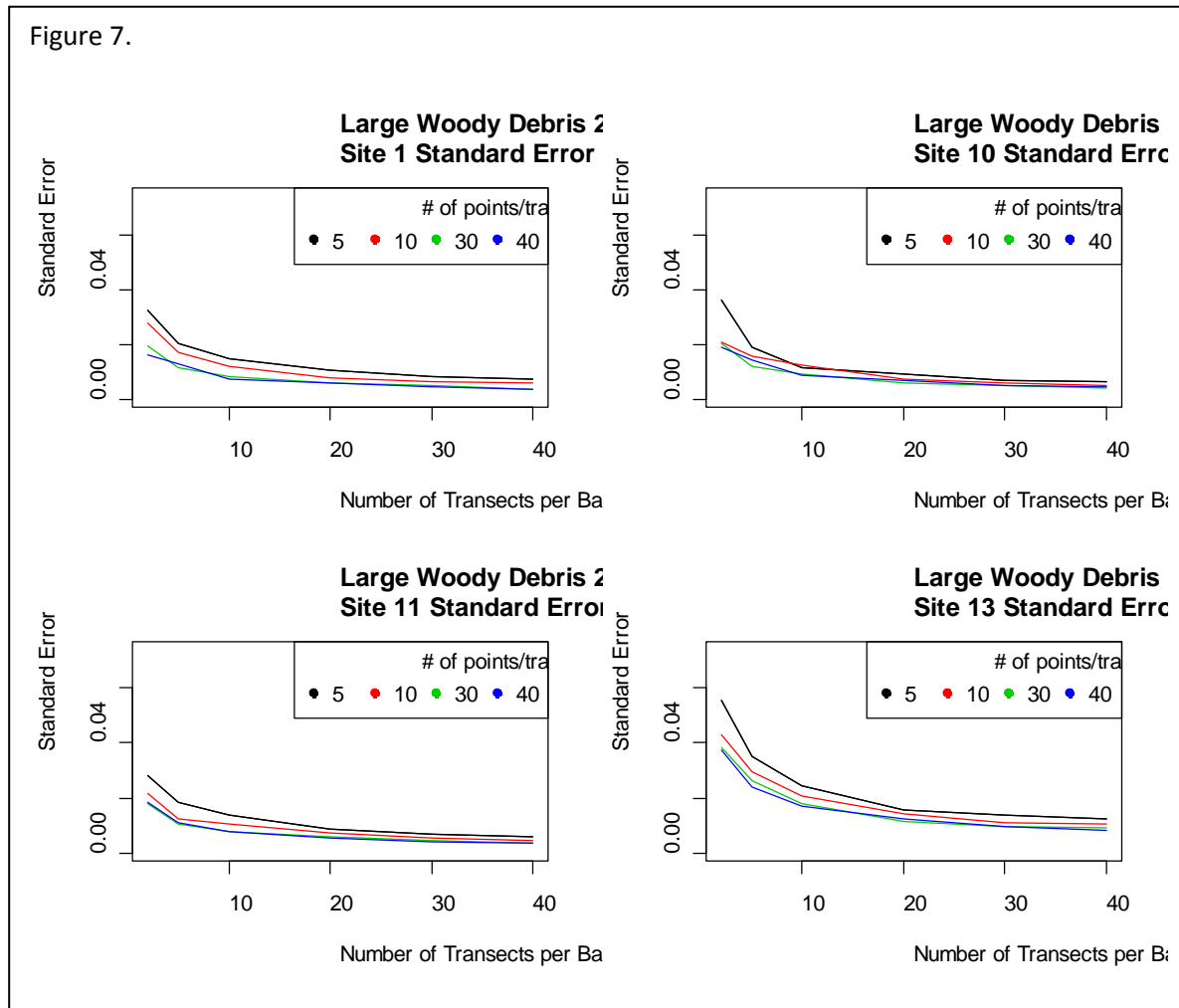
Increasing the number of points per transects results in only modest improvements in reach level estimate precision for Large Woody Debris. For Tap Rooted Perennial, increasing number of points per transect beyond 10 results in nearly no further improvement in reach level precision.



Plots of the reach spatial level standard errors versus number of sites, number of transects, and points per transect, analogous to figures 5 and 6, for all 13 metrics in this analysis are included in Appendix A2. All metrics show similar trends in that beyond 20 or so transects per site, and 10-30 sites points per transect, there is very limited improvement in precision of reach level estimates.

In some cases, interest may lie in estimating a metric at a specific site rather than across an entire GRTS sampling frame. In these cases, it may be important to achieve a desired precision within a site. Figure 7 and 8 show site level precision estimates for Large Woody Debris and Tap Rooted Perennial, by number of transects per banks and number of points per transect, for sites 1, 10, 11, and 13.

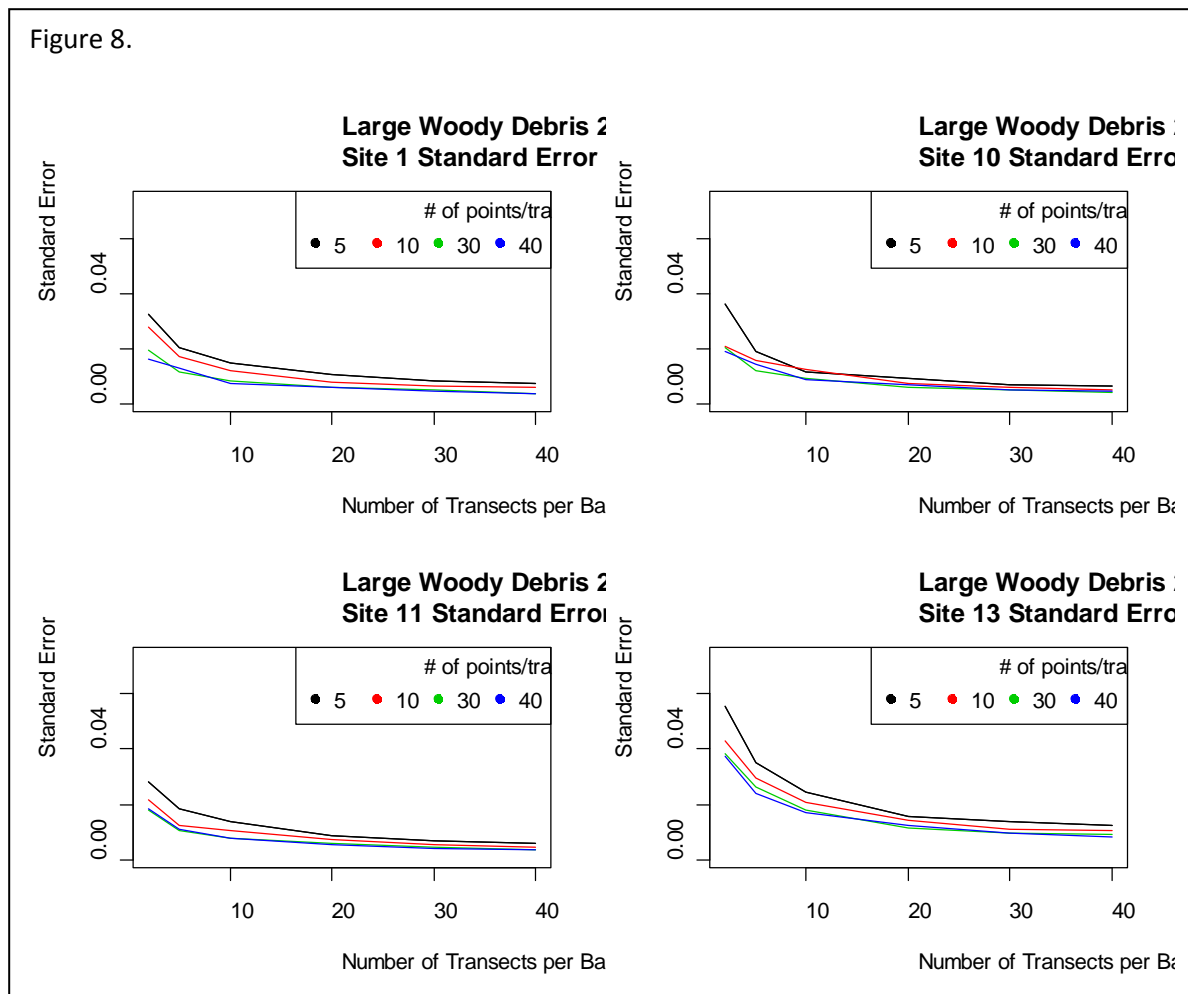
Figure 7.



Because the metrics included in this study are proportions, the variance of each response varies with the magnitude of the response, with highest site level variances occurring as the proportions approach 0.5. Therefore, the site level precision varies by site and metric, even when the same number of transects per banks and points per transect are used. A full set of site level curves showing the tradeoffs between sampling intensity and precision, for all 13 metrics, and all 24 sites included in the 2010 sample, are shown in Appendix A3.

Note that, unlike GRTS sampling frame level estimates, increasing the number of transects will always provide some level of improvement in site level precision estimates.

Figure 8.



Because of the nature of the responses (proportions) cause the variance of the response to vary by the level of the response, some investigation was done to determine whether there may be sampling efficiencies to be gained by optimizing sampling intensity for each site, so that sites likely to have high variability (based on prior estimates) are sampled with higher intensity. However, for a given metric, any potential improvements in efficiency were small. Also, even if there were significant improvements in efficiency to be had for a given metric, when considering multiple metrics there are not consistently high variance sites and low variance sites, but rather the variance tends to be high for some metrics and low for others within each site. Thus, varying sampling intensity by site seems to have little value at this point.

If, however, there is high interest in specific sites, there is certainly no harm to be done by increasing sampling intensity for a given site or sites. The integrity of the GRTS sampling design and resulting reach level estimate will not be negatively impacted by increasing sampling intensity at one or more sites, while keeping sampling intensity the same at all other sites.

Note on the Impact of Changing Sampling Intensities Mid-Study

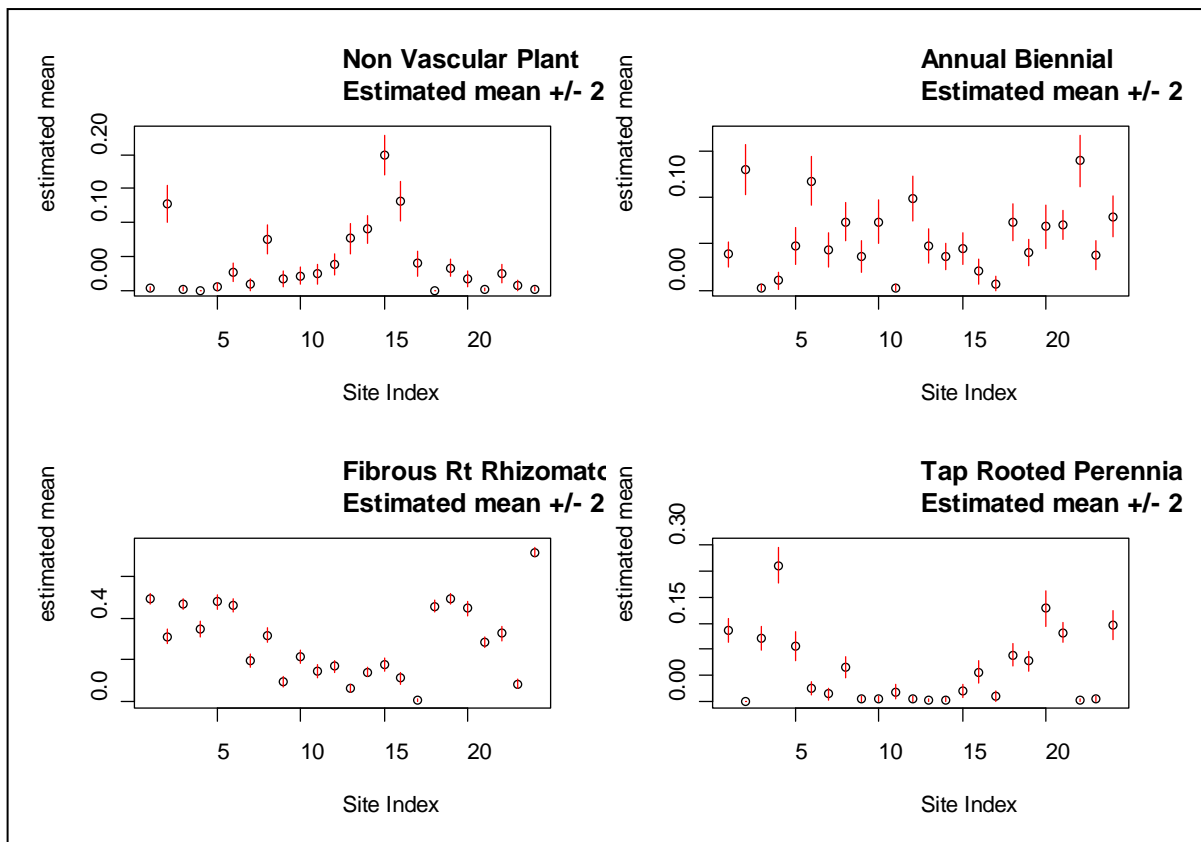
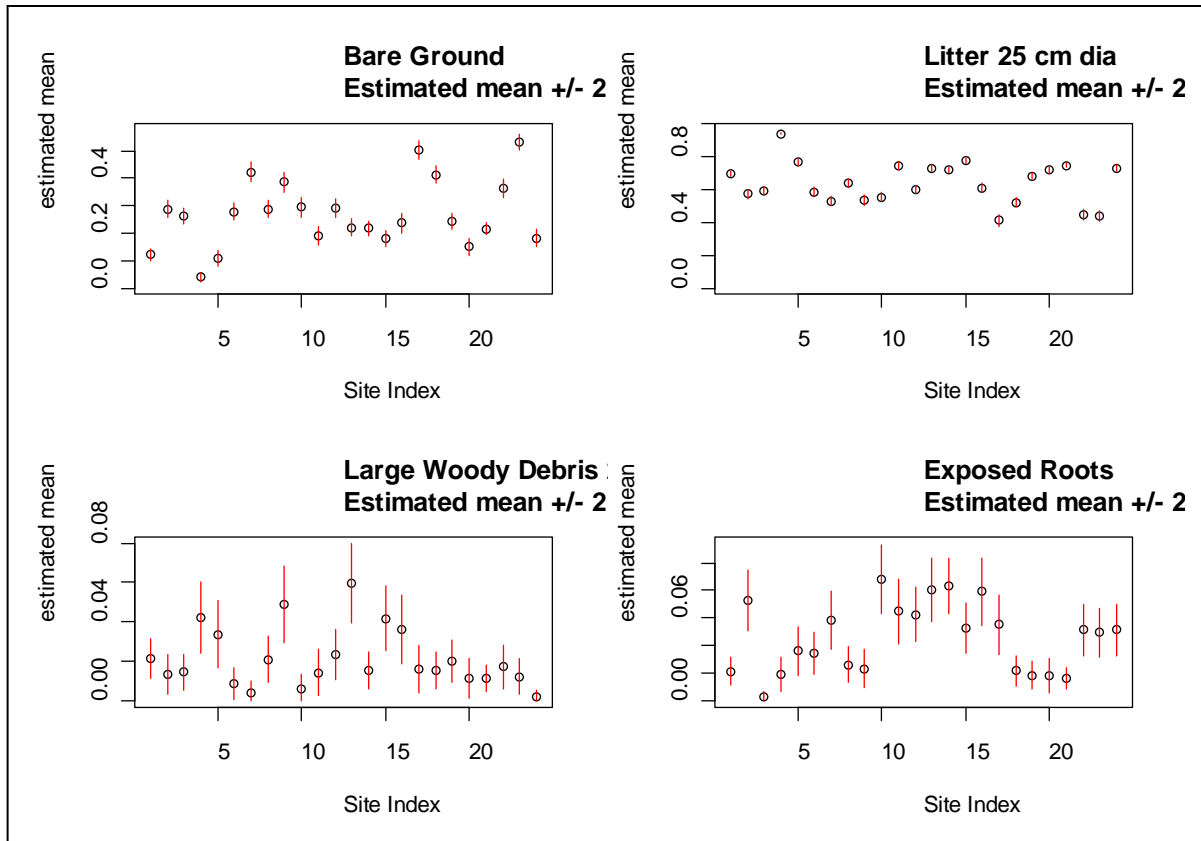
Given that this study applies to long term sampling, and for which sampling has already been completed at all sites for 2010, the question has also arisen as to whether or not changing sampling intensity in proceeding years will adversely affect the integrity of the GRTS design or otherwise bias results. In general, the GRTS design integrity will not be affected, and results will not be biased by changing sampling intensity year to year. In the GRTS analysis, a single response from each site is used. To track trends over time, this response may be, for example, a linear fit of site level response by year. While changing the precision of the estimated response will affect the precision of the linear fit, it will not bias the response, thus it will not bias the GRTS analysis results. If the response of interest is instead, for example, the mean response averaged over multiple years, the analysis will require that the within site level estimate are not biased towards years with more data if changes in sampling intensity are made; but again, as long as the site level response is estimated correctly (which can be done even if sampling intensity changes from one year to another by, for example, giving equal weight to each year), then the integrity of the GRTS analysis is maintained.

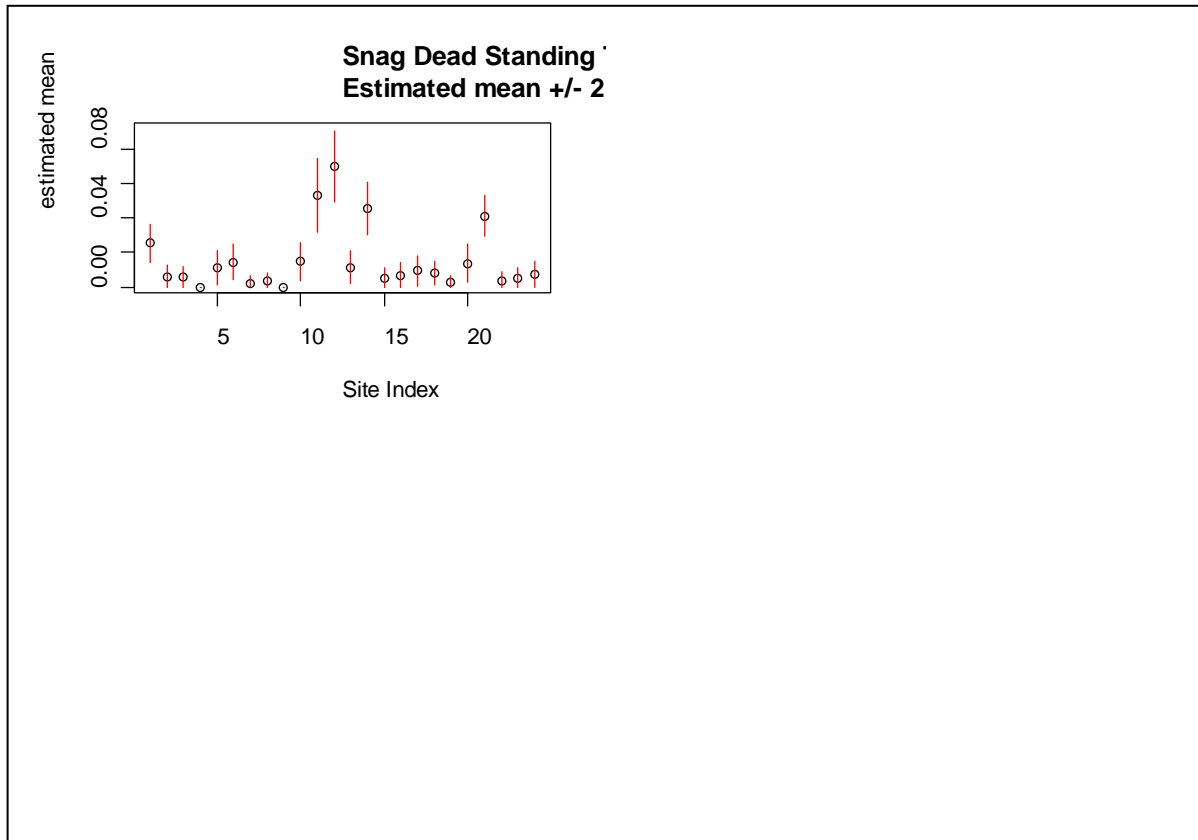
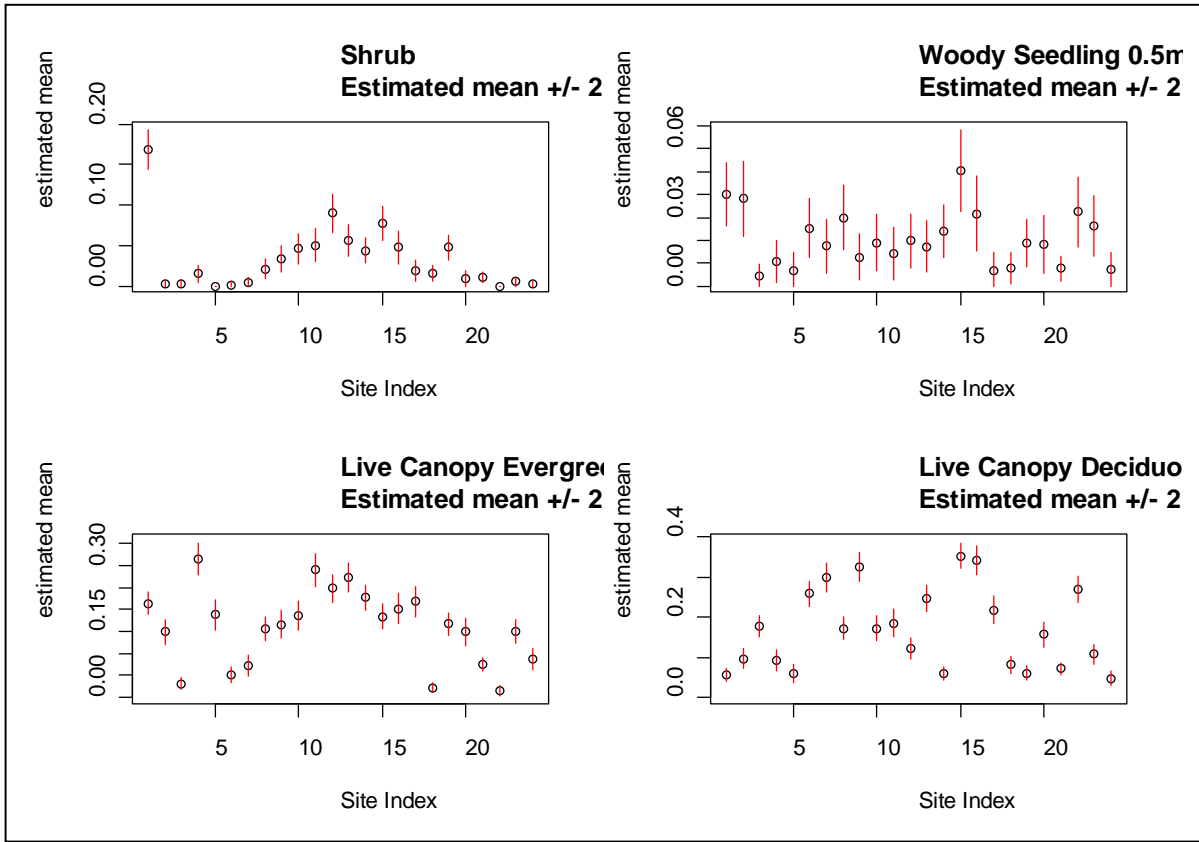
Conclusion

Analysis and simulation from 2010 Sien Merced Riverbank Study data demonstrate that, where reach level estimate are desired, there are within site sampling intensities beyond which increases in sampling intensity provide little additional benefit. Typically, obtaining measurements at more than 20 transects per bank, or more than 10 points per transect, yield little or no benefit in the precision of reach level estimates. Changes in the total number of sites sampled will always yield a change in precision of reach level estimates, with the standard error of reach level mean estimates at least approximately linear with $1/\sqrt{N}$, where N is the number of sites sampled.

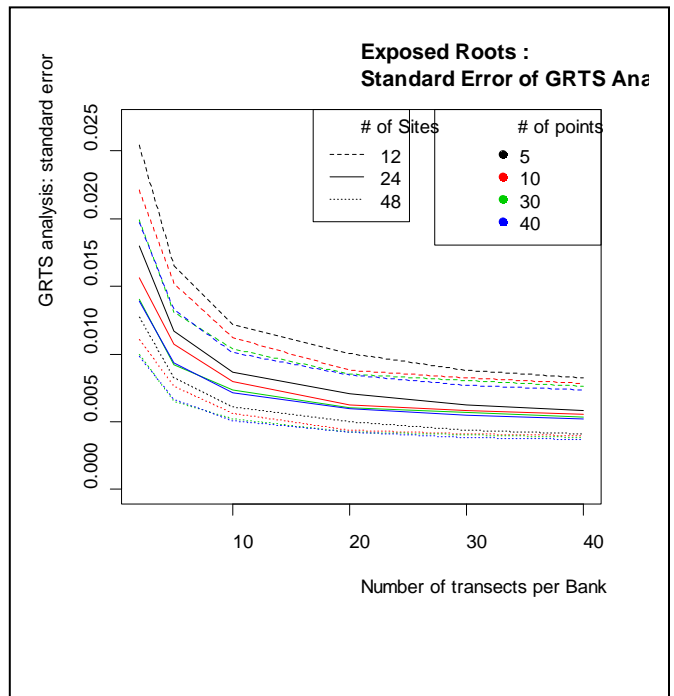
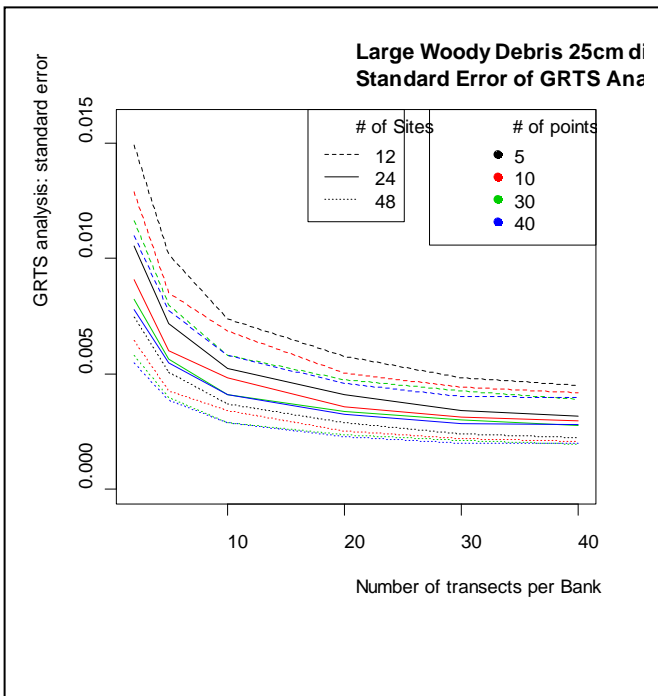
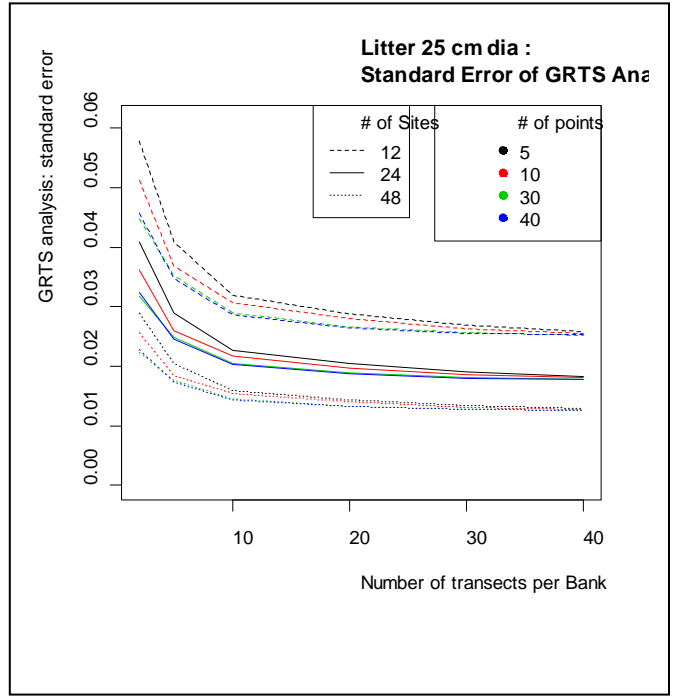
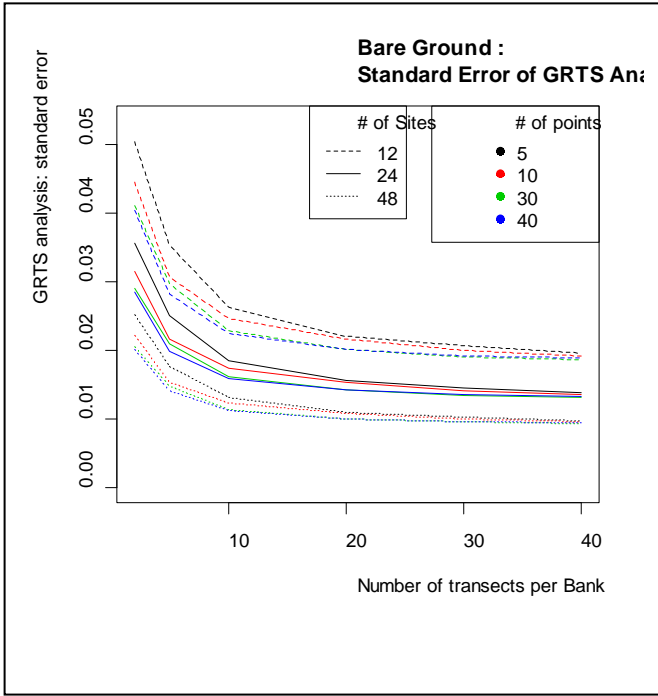
If specific site level estimates are of interest, the researcher may use the curves included to determine a sampling intensity that's likely to achieve the desired precision.

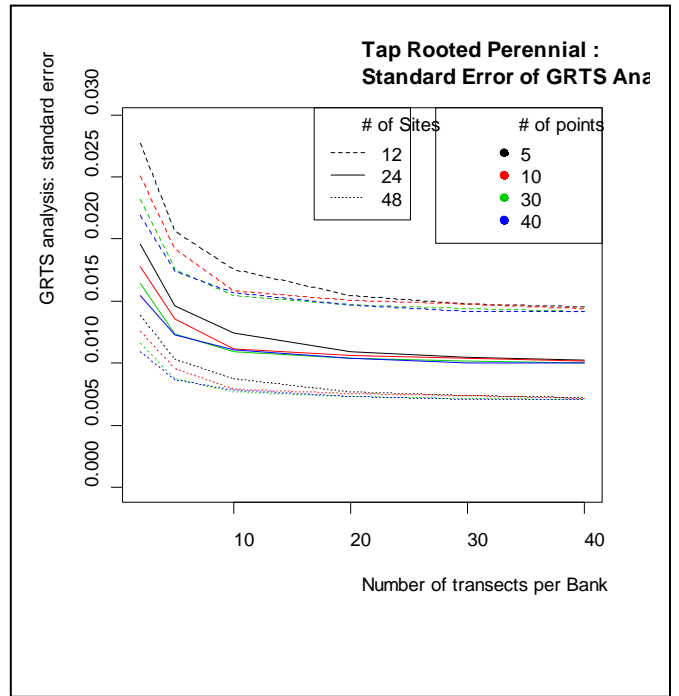
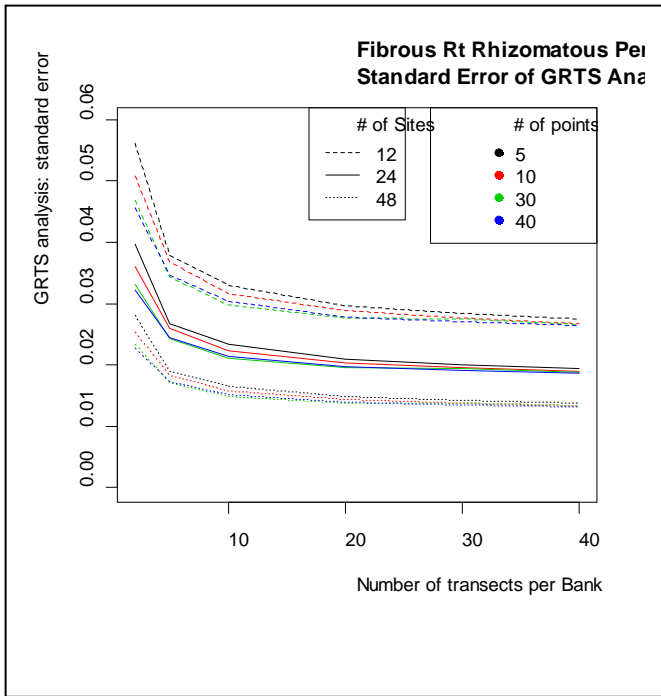
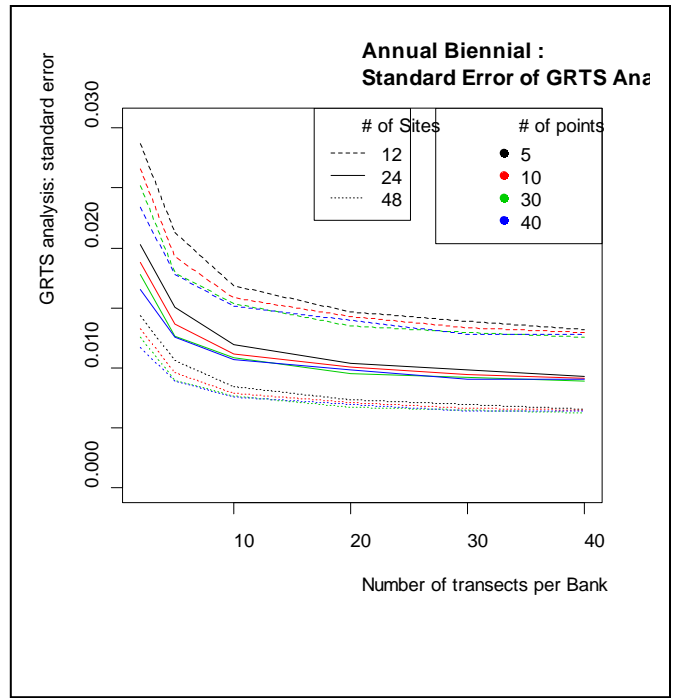
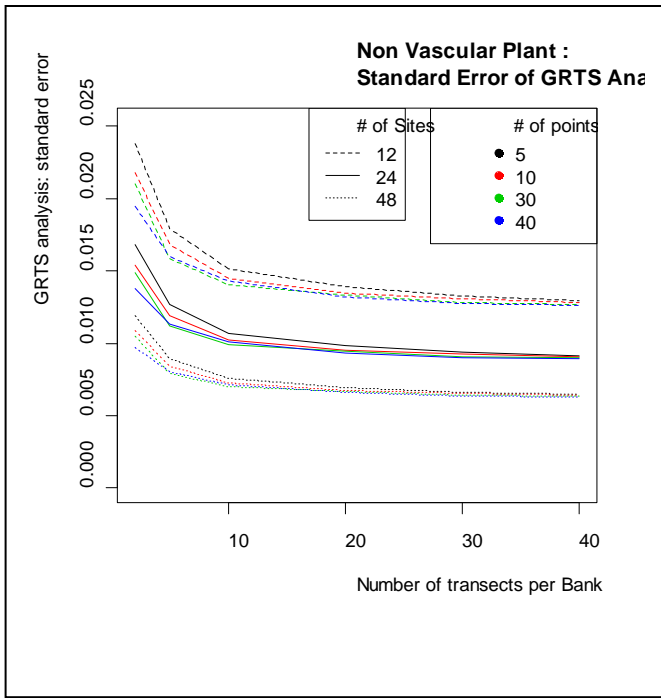
Appendix A1: Estimated Response Average by Site for All Metrics

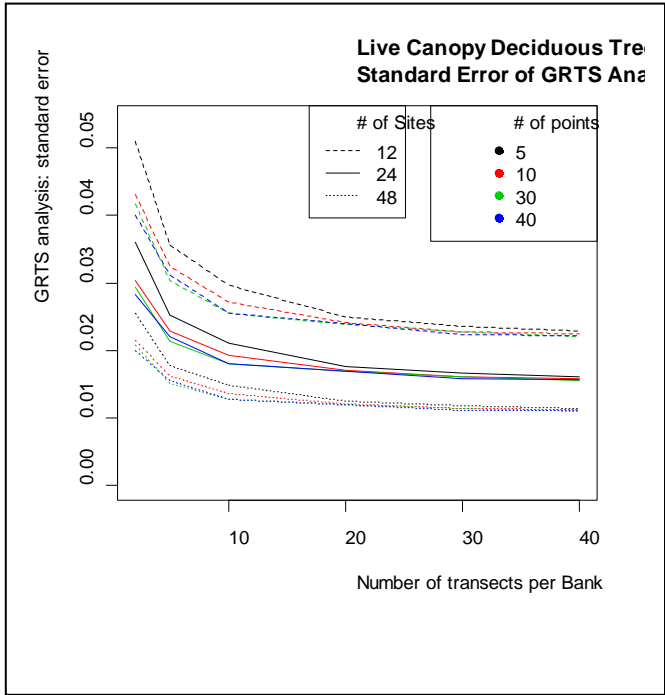
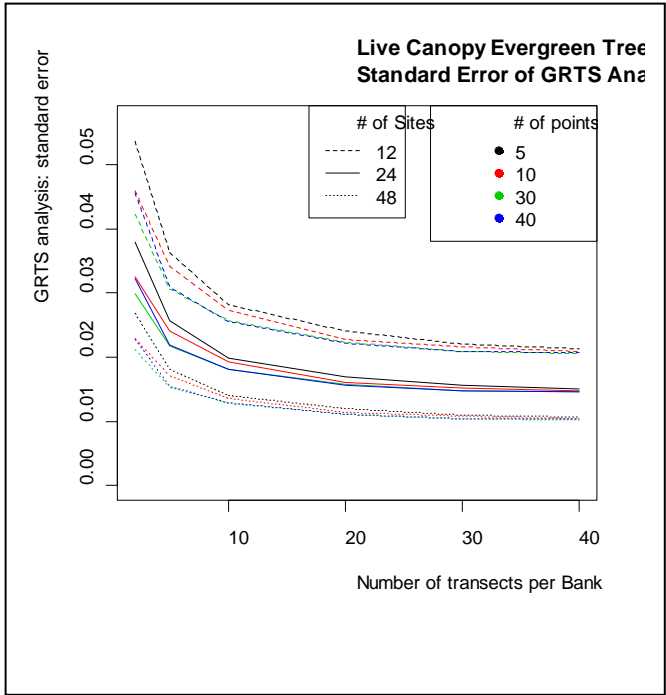
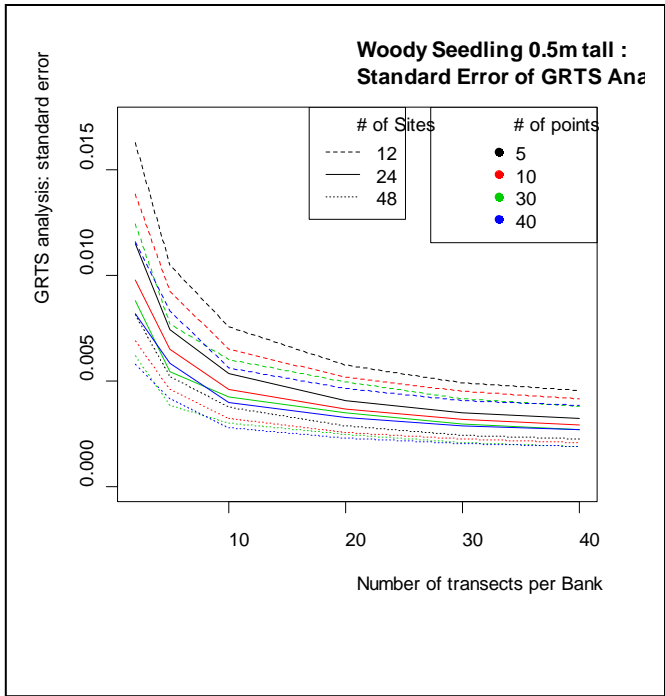
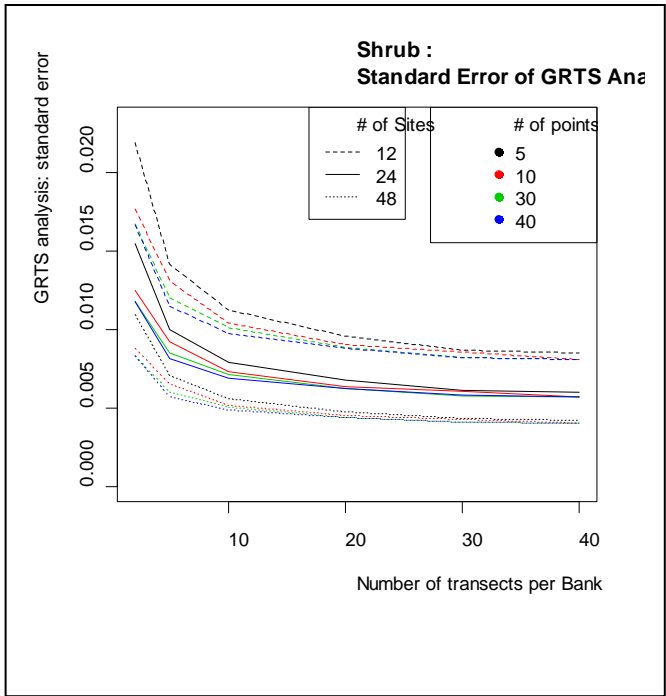


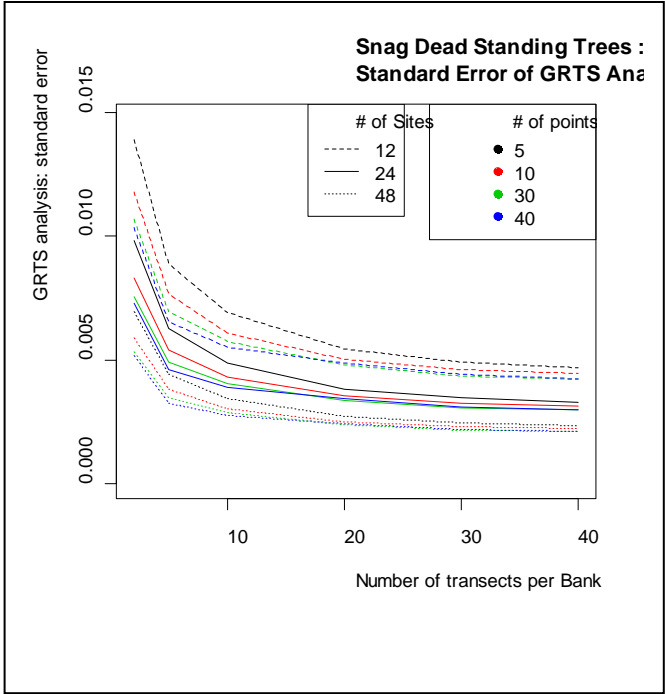


Appendix A2: Site Level Sampling Intensity versus GRTS Sampling Frame Standard Error, for All Metrics









Appendix A3: Site Level Sampling Intensity vs. Site Level Standard Error, for All Metrics and All Sites

