Section 1- Description of methods used to evaluate species richness and other aquatic plant metrics

Section 1.1- Background

A series of White Papers have been developed to evaluate aquatic plant survey data collected from several New York state aquatic plant monitoring programs over the last 100 years. These data can be analyzed from the perspective of several measures used to evaluate aquatic plant communities. Section 2 of this White Paper describes the methods used to evaluate species richness, including projected species richness and suggested species richness scores. The same tools can also be used to evaluate projected mean coefficients of conservatism (or mean C values). **Most of the discussion of aquatic plant survey analysis tools used in this White Paper) and subsequently in White Papers 1D, 1F and 1G, focus on the development of these tools for evaluating species richness. However, as noted below, the same tools can be used for evaluating and calculating mean C values.**

Section 3 outlines tools used in the evaluation and computation of coefficients of conservatism (C values) but NOT used in the evaluation and computation of species richness.

Section 2- Tools Used in the Evaluation of Species Richness (White Paper 1D) and Coefficients of Conservatism (White Paper 1F)

Section 2.1- Background

White Paper 1D provides a detailed summary of species richness in New York state lakes, including a summary of the methods outlined in this White Paper. In addition, White Paper 1F outlines how these methods are used to evaluate coefficients of conservatism, or C values.

Species richness is a common measure of aquatic plant diversity and ecological health. The relationship between species richness and several factors is explored at length in White Paper 1D. These factors include number and density of survey sites, lake area, littoral area, trophic state, latitude, public access, presence (and dominance) of invasive species, management, and annual variability. To evaluate these factors, the raw and/or summarized plant survey data from three of the aquatic plant monitoring programs summarized in White Paper 1A (the NYS BioSurvey lakes, the PIRTRAM lakes, and the AWI lakes; the ALSC data cannot be used to calculate species richness due to the lack of species-level identifications) need some manipulation, or may be dependent on the definition of several specific terms. These are discussed in Sections 2.2 to 2.8.

The same applies to coefficients of conservatism, a common measure of the quality of individual plant species found during aquatic plant surveys.

Section 2.2- Aquatic Plant Survey Sites

The NYS BioSurvey and the ALSC studies provide species lists (BioSurvey) or genera lists (ALSC) for all of the surveyed lakes, but the extent of the surveying is not provided in the

remaining documentation from those programs. While it is assumed that the entire littoral zone was surveyed in all NYS BioSurvey and ALSC lakes, the number and location of the aquatic plant survey sites are not available for the surveyed lakes, and therefore cannot be used in any analysis of species richness (or other plant community metrics) relative to surveying effort. In addition, as introduced in Section 2.3 below and as discussed in White Paper 1D, projected species richness- the number of unique species if a standardized survey site density was used- cannot be accurately estimated without individual survey site data. However, species richness for the NYS BioSurvey lakes can be calculated

For each of the approximately 200 surveys on more than 50 PIRTRAM lakes, the individual aquatic plants associated with point-intercept sites surveyed using a combination of two-sided rake tosses and visual assessments were documented. For a subset of these lakes, discussed below, "granular" point-intercept data are available, documenting the presence and/or relative abundance of all plants encountered at each surveyed site; the absence of a documented presence is assumed to represent the absence of that plant at that site. It is also assumed that these point-intercept sites are equally distributed throughout the littoral zone. The granular survey site data available for many of the PIRTRAM lakes includes multiple data assessments, including projected species richness and the number of survey sites needed to estimate species richness.

In contrast, the AWI surveys used a combination of rake tosses (albeit not conducted in traditional point-intercept sites within overlay grids) and serpentine bed surveys to generate less granular data. Single plant abundance "scores" (trace, sparse, moderate or dense) are assigned to each plant bed, without any indication of plant frequency within these beds. Similar to the relative abundance scores assigned to entire NYS BioSurvey lakes, these assigned values require a presumption that this abundance score is uniform throughout the plant bed. This limits the use of these data, although as seen below, the relative abundance of individual plants (including AIS species) can be evaluating using this blended rake toss/visual bed assessment data.

Section 2.3- Standardized Projected Survey Sites

Any comparison among the monitoring programs discussed in White Paper 1A may be adversely affected by inconsistencies in the number of survey sites used between programs, and even in some lakes (and some lake-years) within programs. This potential discrepancy between actual numbers of survey sites and any recommended density of survey sites is not relevant for the NYS BioSurvey and ALSC programs, since individual survey sites (overall number, distribution, or density) data are not available for those lakes, and perhaps not for the AWI lakes due to the blend of granular rake toss data and broad summary of relative abundance for entire plant beds. However, to facilitate a comparison among PIRTRAM lakes, a standardized number of survey sites, based on available survey area, should be defined. A casual inspection of the PIRTRAM survey lakes might lead to assumptions that all surveys were

Table 2.3- Plant Survey Site Densities in PIRTRAM Lakes

	PIRTRAM Density
Minimum	0.2 sites per ha
25th Percentile	0.8 sites per ha
50th Percentile	1.4 sites per ha
75th Percentile	3.7 sites per ha
Maximum	13 sites per ha
N Lakes	50 lakes, 165 lake years^

^ lake year includes each year a lake sampled

conducted in the same way. While all surveys appear to involve rake toss collections from sites associated with overlay grids, the number and density of survey sites often varied significantly from lake to lake. Many of the surveys on waterbodies conducted in support of the NYSDEC enhanced review program used grid sizes that met or exceeded the 1 site per hectare requirement, while many of the other surveys- conducted by lake associations or NYSDEC not covered by the enhanced review requirements- included a wider range of site densities. The actual density of PIRTRAM survey sites in each lake is provided in White Paper 1A, Appendices 1.1 and 1.2.

A summary of the site densities in the PIRTRAM lakes is provided in Table 2.3. The range of survey site densities in the PIRTRAM lakes dataset spans from about 13 sites per hectare of littoral area (Lake Rippowam) to 0.2 site per hectare or about 5 hectares per site (Chautauqua Lake and Kinderhook Lake), or a difference of a factor of about 25. However, survey site densities in most PIRTRAM lakes fall within a much narrower range. In general, most of the PIRTRAM lakes were surveyed at the rate (1 site per hectare) recommended by the NYSDEC in the previous Enhanced Review process, as discussed in White Paper 1B.

As noted above, while the maximum number of plant taxa (i.e. maximum or projected species richness) in each lake is essentially unbounded (or at least theoretically bound only by the limits of complete surveys of the entire littoral zone), a practical upper limit for species richness needs to be defined. This is also apparent from logarithmic descriptions of the plant taxa distributions described above, since asymptotes are only achieved with extremely large numbers of survey points. However, existing plant surveys in smaller lakes (Oscaleta Lake), and extremely refined lake surveys looking for rare invasives (Cayuga Lake re: hydrilla) or protected plants (Lake Luzerne) suggest a tight (high density) practical upper limit of aquatic plant survey density was achievable through replicated monitoring conducted by Racine-Johnson Aquatic Ecologists and Allied Biological/DFWI, respectively.

These studies roughly corresponded to a practical upper limit of 4 sites per littoral acre, which can be used to estimate the maximum species richness, as discussed in White Paper 1D (Species Richness). This also roughly corresponds to 30m x 30m grids, which itself corresponds to adjacent 10-15m rake toss "reach" grids with a small buffer (3-5m) between grids. In other words, 4 sites per acre is close to the maximum number of sites that would avoid overlapping grids when the two-sided tethered rakes are tossed the recommended distance as per PIRTRAM protocols. However, a survey with such as tight density of aquatic plant surveyed sites may not be practical, particularly in very large lakes. A less dense survey site density may be more practical (and therefore more achievable given available resources), less vulnerable to extrapolation errors when projected species richness (or C values) are required (see below), and is more likely to be consistent with historical or existing state survey requirements. As discussed more in White Paper 1D, a standardized survey site density of 1 site per littoral hectare, consistent with the historical NYDEC monitoring requirements, should be achievable in most large and small lake surveys, will likely find most aquatic plant species in a lake, and allows for consistent comparison across lakes. Unless otherwise noted, the ***projected number of aquatic plant survey sites (and associated species richness and mean C values given this survey site density) corresponds to a 1 site per littoral hectare site density.***

Section 2.4- Observed Species Richness and C Values

Observed species richness is simply the number of unique aquatic plant taxa found during a survey. As noted above, this is generally limited to submergent plants, floating leaf plants, and those emergent plant species growing at or below the water surface at the time of the survey (and therefore most likely to be characterized as submergent or floating). An example of the latter is *Polygonum amphibian* (water smartweed), likely the only *Polygonum* species to be encountered as a submergent or floating leaf plant. Given the challenges in accurately identifying some plant taxa due to maturity of the plant, lack of collected distinguishing flowers, leaves, and reproductive organs, plasticity, and other factors cited above, some "species" are only reported to genera. Examples of this include *Sparganium* (bur reed), *Nuphar* (yellow water lily), and *Nitella* (stonewort). Although these and other identified plants are genera rather than species, they were documented and counted here as unique species and therefore are included in species richness counts. This is further discussed in White Paper 1B.

The observed species richness values provided in this White Paper represent the unique aquatic plant "species" counts associated with the number of sites surveyed. However, as discussed at length in the section related to Projected Species Richness (Section 2.5), the estimated number of unique aquatic plant species associated with various intervals of aquatic plant survey sites can be used to calculate an estimated Projected Species Richness, which represents a standardized means for evaluating species richness.

The same logic applies to computation and evaluation of coefficients of conservatism. Mean C values for a surveyed lake may be dependent on several factors, and most importantly would likely vary depending on the number of survey sites. As with species richness, a standardized projected mean C value will allow for comparison across lakes (and improve comparison of floristic quality indices, which, as seen in White Papers 1D, 1F and 1G, include both species richness and mean C values.

Section 2.5- Projected Species Richness and Mean C Values

As noted above, the observed species richness is an aquatic plant survey may be a function of the number of surveyed sites- this issue is explored at length in White Paper 1D. To facilitate a comparison of species richness across multiple programs (each of which may have specific criteria for establishing survey site densities), a standardized survey site density should be established. In the Section 2.3 discussion above, a standardized site density of 4 sites per littoral acre is recommended, although it might not be achievable in most surveys. Whether this survey site density is operationally deployed or used for an endpoint to "project" species richness, adopting this survey site density will allow for a common platform for reviewing species richness data.

The expected number of aquatic plant taxa (i.e. species richness) in a given number of aquatic plant survey sites can be generated using the highly granular data in many of the PIRTRAM study lakes. This requires an estimate of cumulative plant taxa found in combinations of plant survey sites, which can be used both to estimate species richness at increments of plant surveys, and to extrapolate maximum species richness with more sites than were actually sampled. For example, relationships between the expected number of aquatic plant taxa and various intervals

of plant survey site numbers can be used to estimate the maximum species richness at a defined higher number of survey sites.

The process for evaluating projected species richness as a function of a high standardized number of survey sites (= 4 sites per littoral acre) is described below using examples of PIRTRAM lakes. **Evaluations of projected species richness for the PIRTRAM lakes data are discussed at length in White Paper 1D.**

Section 2.5.1- Subsampling methods

These cumulative taxa estimates require data subsampling given the extraordinarily large number of site combinations that exist for these surveys. For example, in Cazenovia Lake (a PIRTRAM lake), there are 304 combinations of survey sites if a single survey site is evaluated, given 304 survey sites used in the surveys conducted by Racine Johnson Aquatic Ecologists from the late 2000s to 2019. However, when two sites are evaluated to estimate species richness, there are more than 46,000 combinations of 304 sites available, based on the formula in Equation 2.4.1:

> *Equation 2.5.1: nCr = n!/(o! x (n-o)!), when n = total number of sites and o = combination of sites*

In this case, nCr = 304!/(2! × (304 - 2)!)), and when 150 sites are evaluated, more than 10^{60} combinations of sites are possible. Even with a smaller lake with fewer sampling sites, such as Creamery Pond with 21 survey sites, evaluating two sites offers 210 unique combinations of survey sites. Since the evaluation of all sites is both prohibitively time consuming and beyond the capabilities of most coding scripts, an alternative approach to estimating species richness is needed.

Smith et. al (1995) summarized a bootstrapping analysis for estimating mean cumulative (bird) species counts using all combinations of six survey points and six visits, as per resampling methods outlined by Efron (1982). Given the very large number of aquatic plant survey points in many New York state lakes, and need for extensive resampling to reduce expected variance between combinations of sampling sites, an alternative method for estimating cumulative means was devised for this study. This is referred to below as a 'modified bootstrap analysis'.

Section 2.5.2-Modified bootstrap analysis

A modified bootstrap analysis is summarized below for estimating cumulative species richness at various intervals of survey sites. For each of the PIRTRAM lakes with granular survey data, survey sites were numbered consecutively, with only littoral zone survey sites chosen for this study. Cumulative mean species richness was estimated using (up to) 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 100, 125, 150, 175, 200, 250 and 300 survey sites, with the number of evaluated sites for each lake chosen as the largest number less than (or equal to) the number of survey sites. So, using the example above, mean cumulative (estimated) species richness in Creamery Pond (with 21 survey sites) was evaluated using 1, 2, 3, 4, 5, 10, 15 and 20 survey sites. Survey site combinations were chosen by random number generations, with care taken to avoid replicate combinations for any "resampling" event. Mean cumulative species richness was calculated in MS Excel as follows (it is likely that other statistical tools, including R programming, could also be used to achieve the same objectives). For each group of survey sites,

the first site (row) was chosen by random number generation, with all subsequent survey sites (rows 2 through X) assigned a different random number using the array formula:

Equation 2.5.2.1
*=LARGE(ROW($1:$X)*NOT(COUNTIF(C$2:C2,ROW($1:$X))),RANDBETWEEN(1,(51+2-1)-ROW(C2))),*
with X corresponding to the total number of lake survey sites.

This process assured no duplicate survey site numbers in each group of sites.

The average number of plant taxa in each group of survey sites (hereafter referred to as a "run") was measured using the following array formula:

Equation 2.5.2.2
=IF(YrLakeB!$KT5=0,"",MAX(IF((YrLakeB!$B$1:$KS$1=B$2:B$X),YrLakeB!$B5:$KS5)));

This formula excludes from the cumulative species count any plants not observed at any survey site included in the site combinations, and counts the most abundant survey result (using the abundance scale described earlier in this report) in each combination as documented in the MS Excel sheet labeled YrLakeB. The latter corresponds to an S x T matrix of S survey sites at which the number of plant taxa T were documented (as species presence or relative abundance) during these surveys.

The expected cumulative number of plant taxa can be calculated given varying intervals of aquatic plant survey sites- i.e. expected species richness found in 5 sites, 10 sites, 50 sites, and so on.

Section 2.5.3-Variance analyses to determine the optimal number of runs or sites

Section 2.5.3.1- Background

The variance associated with these results decreases significantly as both the number of "runs" (combinations of sites using random number generation of individual site numbers) increases AND as the number of sites increases. Resampling was conducted in intervals of 25 random number generations ("runs") up to 100 runs for smaller lakes and 200 runs for larger lakes (so each combination of survey sites in Creamery Pond was evaluated in runs of 25, 50, 75 and 100 combinations of survey sites, corresponding to 25, 50, 75 and 100 different and unique combinations of the 21 survey sites for the lake). As noted above, for each set of runs, random numbers were independently generated to avoid site numbers in larger runs including previously generated (smaller) site number runs. Cumulative mean and standard deviations were calculated for each set of runs.

To evaluate optimal (most efficient) numbers of survey sites and the optimal (reduced variance relative to the computational effort) number of survey runs, one-way ANOVA analyses were conducted using 95% confidence intervals to determine the variance among each group of survey sites and runs (https://acetabulum.dk/anova.html). Variance was evaluated using the Tukey-Kramer HSD Post-Hoc Test, with each set of survey sites and runs assigned a p value and significance level associated with similarity of these runs relative to all other evaluated runs.

As per standard statistical criteria, it is assumed that a p-value less than 0.05 indicates statistical significance. Sets of (consecutive) survey runs with little to no variance between these sites exhibited very low p-values and significance levels, while high variance between run sets exhibited higher p-values and significance (of difference) levels. It is presumed that the point at which the p-value shifts from >0.05 to <0.05 corresponds to the approximate inflection point of these cumulative species richness values in these survey sites- in other words, this inflection point represents the number of sampling sites required to shift from a high degree of change (in species richness) per unit effort to a low degree in change in species richness. This likely represents the optimal sampling effort above which additions of sampling sites results in only small changes in species richness. **However, this approach also provides value only if correlative relationships between optimal sampling effort and maximum species richness can be established (see White Paper 1D).**

It should also be noted that some outlier data exists, with some combinations of sampling sites yielding much different p-values and significance level than expected compared to slightly higher or lower combinations of sites. For example, in a lake with 100 survey sites, if the inflection point of a p-value of 0.05 appears to occur between 20 and 25 sites, but much higher p values unexpectedly occur between 50 and 60 sites, professional judgement is used to most accurately assign the inflection point.

Section 2.5.3.2- Optimal number of survey sites

As noted above, the ANOVA process can be used to evaluate the optimal number of survey sites to maximize sampling effort (in this case, the point at which adding additional survey sites does not significantly increase the likelihood of finding new unique taxa relative to the additional effort, recognizing that (as discussed in White Paper 1D) the number of new species generally increases with increasing numbers of survey sites). Specifically, the optimal number of survey sites can be calculated from the point at which the p-value shifts from >0.05 to <0.05 corresponding to the approximate inflection point of these sets of survey runs. In other words, this inflection point represents the number of sampling sites required to shift from a high degree of change (in species richness) per unit effort to a low degree in change in species richness. This likely represents the optimal sampling effort above which additions of sampling sites results in only small changes in species richness. This value, the point (number of survey sites) at which additional sampling is unlikely to yield significant increases in species richness, should be higher than the minimal values needed to find the majority of the unique taxa in the lake, or minimal values needed to project species richness. The sampling efforts needed to achieve these goals are discussed at length in White Paper 1E (finding the majority of unique and specific (AIS and/or RTE) taxa, and White Paper 1D (Species Richness), respectively.

Table 2.5.3.1 shows the optimal number of plant survey sites in the PIRTRAM lakes with granular survey data- this indicates the number of sites corresponding to the most efficient process for identifying unique taxa in a lake (any additional sites are unlikely to provide a significant increase in additional unique taxa relative to the effort expended). The data in this table can be summarized as follows:

Table 2.5.3.1- Most Efficient (Optimal) # Sampling Sites to Maximize Sampling Effort, PIRTRAM Lakes

Lake Name	Year	Littoral Area (ha)	#Survey Sites	#Optimal Sites
Ballston Lake	2006	48	34	25
Big Fresh Pond	2006	13	19	>15
Blydenburgh Lake	2012	40	27	15
Blydenburgh Lake	2014	40	27	3
Cazenovia Lake	2010	225	304	40-50
Cazenovia Lake	2011	225	304	50-60
Cazenovia Lake	2012	225	304	50-60
Cazenovia Lake	2013	225	304	20-40
Cazenovia Lake	2014	225	304	40-60
Cazenovia Lake	2015	225	304	40-50
Cazenovia Lake	2016	225	304	50
Cazenovia Lake	2017	225	304	15-20
Cazenovia Lake	2018	225	304	20-25
Cazenovia Lake	2019	225	304	15-20
Collins Lake	2007	5	38	>25
Creamery Pond	2008	4	18	3
Creamery Pond	2009	4	18	>15
Creamery Pond	2010	4	21	10
Creamery Pond	2011	4	21	10
Creamery Pond	2012	4	21	10
Creamery Pond	2013	4	21	>20
Hards Pond	2010	12	18	>15
Hards Pond	2011	12	18	>15
Java Lake	2008	21	16	>15
Java Lake	2009	21	16	10
Java Lake	2010	21	16	>15
Kinderhook Lake	2006	109	20	10
Kinderhook Lake	2007	109	20	4
Lake Luzerne	2010	24	168	>150
Lake Ronkonkoma	2009	21	22	>20
Lake Ronkonkoma	2010	21	22	>20
Lake Ronkonkoma	2011	21	22	>20
Lake Ronkonkoma	2012	21	22	>20
Lake Ronkonkoma	2014	21	22	15
Lamoka Lake	2006	166	180	25
Lamoka Lake	2008	166	180	>180
Lamoka Lake	2009	166	180	50-60
Morehouse Lake	2010	35	30	>30
Quaker Lake	2010	64	30	15
Saratoga Lake	2010	657	241	125
Saratoga Lake	2011	657	304	>300
Saratoga Lake	2012	657	304	>300
Snyders Lake	2002	15	40	15
Snyders Lake	2003	15	48	20
Snyders Lake	2004	15	57	15
Snyders Lake	2005	15	32	10
Snyders Lake	2006	15	40	15
Snyders Lake	2007	15	57	20
Snyders Lake	2008	15	57	25
Snyders Lake	2009	15	55	15
Snyders Lake	2010	15	44	25-30
Snyders Lake	2011	15	51	20
Waneta Lake	2006	170	146	>125
Waneta Lake	2008	170	146	50
Waneta Lake	2009	170	146	>125

a. One measure of the optimal number of survey sites, the sites required to maximize the surveying effort relative to identifying unique taxa within each lake, cannot be evaluated in many lakes with few total survey sites, even if the density of these survey sites (# sites per littoral area) is relatively large. In these cases (for example, Hards Pond, Java Lake, Lake Ronkonkoma, etc.), a more refined (denser site distribution) surveying matrix would be required to determine optimal sampling effort. This also occurs in some larger lakes with far more surveying sites (for example, Lake Luzerne, Waneta Lake in 2006 and 2008….); for these lakes, even more survey sites would continue to provide more unique (cumulative mean) taxa commensurate with the additional sampling effort. For these lakes, the optimal number of sampling sites cannot be accurately estimated from the existing number of survey sites. However, in general optimal sampling effort can be estimated, at least using the methods described above, in most large and small lakes.

b. It is possible that inflection points exist in graphical summaries of these data, but these inflections- a shift from lesser to greater concavity- do not appear to be statistically significant (cannot be ascertained through the aforementioned ANOVA analysis). This indicates that additions of survey sites, throughout the entire range surveyed for these lakes, do not yield results (cumulative number of unique plant taxa) that are significantly different. For these lakes, it is likely that far more survey sites would be required to find a statistically significant inflection point indicating an optimal number of survey sites (indicating a statistically significant decrease in results relative to the effort associated with adding survey sites).

For many smaller lakes that already have high densities of survey sites per littoral area, the approach outlined above for identifying optimal sampling does not seem to converge on a recommended sampling site frequency. These discrepancies suggest that the **number of surveying sites on any sampled lake should be based on factors other than optimizing sampling effort (such as based on approximating maximum species**

richness, finding AIS, or other factors). These are discussed in Section 2.4 and other White Papers in this series

c. A second measure of the optimal number of survey sites, defined as the minimum number of survey sites (as calculated using ANOVA) to optimize sampling effort relative to measurement of species richness, varies significantly between lakes and within lake survey seasons. In general, the number of optimal sampling sites, as defined above, was smaller in small lakes (with a smaller number of total taxa) than in larger lakes, but these differences appeared to be much smaller than both the intra- and inter-lake variability. For example, the optimal number of survey sites in Cazenovia Lake over a ten year period varied from 15 to 60 sites, despite a roughly similar distribution of plant species from year to year. This wide variability also occurred in mid-sized lakes with smaller number of survey sites (Snyders Lake, varying over ten years from 10-25 optimal sites) and small lakes with even fewer sampling sites (Creamery Pond, varying over six years from 3 to more than 20 sites needed to optimize surveying). The differences in optimal numbers of survey sites from lake to lake does not appear to be strongly influenced by the number of survey sites per littoral area or the number of invasive plants (which presumably could depress overall species richness).

Table 2.5.3.2- Survey Runs to Stabilize Species Richness on PIRTRAM lakes with granular survey data

Lake Name	Year	Littoral Area (ha)	Lake Area (ha)	#Survey Sites	#Runs Optimal Stabilized
Ballston Lake	2006	48	107	34	50-75
Big Fresh Pond	2006	13	34	19	25
Blydenburgh Lake	2012	40	40	27	75-100+
Blydenburgh Lake	2014	40	40	27	50-75
Cazenovia Lake	2010	225	471	304	100-125
Cazenovia Lake	2011	225	471	304	100-125
Cazenovia Lake	2012	225	471	304	125-150
Cazenovia Lake	2013	225	471	304	150-175
Cazenovia Lake	2014	225	471	304	100-125
Cazenovia Lake	2015	225	471	304	150-175
Cazenovia Lake	2016	225	471	304	125-150
Cazenovia Lake	2017	225	471	304	150-175
Cazenovia Lake	2018	225	471	304	150-175
Cazenovia Lake	2019	225	471	304	150-175
Collins Lake	2007	5	23	38	50-75
Creamery Pond	2008	4	4	18	50-75
Creamery Pond	2009	4	4	18	75-100+
Creamery Pond	2010	4	4	21	25-50
Creamery Pond	2011	4	4	21	50-75
Creamery Pond	2012	4	4	21	75-100+
Creamery Pond	2013	4	4	21	25-50
Hards Pond	2010	12	12	18	25
Hards Pond	2011	12	12	18	25-50
Java Lake	2008	21	21	16	25-50
Java Lake	2009	21	21	16	25-50
Java Lake	2010	21	21	16	25
Kinderhook Lake	2006	109	138	20	25-50
Kinderhook Lake	2007	109	138	20	75-100+
Lake Luzerne	2010	24	40	168	50-75
Lake Ronkonkoma	2009	21	92	22	75-100+
Lake Ronkonkoma	2010	21	92	22	25-50
Lake Ronkonkoma	2011	21	92	22	50-75
Lake Ronkonkoma	2012	21	92	22	50-75
Lake Ronkonkoma	2014	21	92	22	25-50
Lamoka Lake	2006	166	294	180	125-150
Lamoka Lake	2008	166	294	180	100-125
Lamoka Lake	2009	166	294	180	100-125
Morehouse Lake	2010	35	43	30	50-75
Quaker Lake	2010	64	112	30	50-75
Saratoga Lake	2010	657	1526	241	150-175
Saratoga Lake	2011	657	1526	304	75-100
Saratoga Lake	2012	657	1526	304	175-200+
Snyders Lake	2002	15	45	40	50-75
Snyders Lake	2003	15	45	48	75-100+
Snyders Lake	2004	15	45	57	25-50
Snyders Lake	2005	15	45	32	50-75
Snyders Lake	2006	15	45	40	25-50
Snyders Lake	2007	15	45	57	50-75
Snyders Lake	2008	15	45	57	50-75
Snyders Lake	2009	15	45	55	25
Snyders Lake	2010	15	45	44	75-100+
Snyders Lake	2011	15	45	51	50-75
Waneta Lake	2006	170	317	146	100-125
Waneta Lake	2008	170	317	146	125-150
Waneta Lake	2009	170	317	146	125-150

d. These findings suggest that this method for evaluating optimal number of plant survey sites is not predictable, and should not be used *a priori* to define sampling sites. As discussed below, the choice of the number of plant survey sites most appropriate for a monitoring program should be based on other methods. While this discussion should account for the optimal number of plant survey sites needed to maximize sampling effort, other factors should also be

considered. **NOTE that these other factors- the need to find all AIS or identify all species in the lake (i.e. projected species richness), permitting requirements, minimizing surveying while still projecting species richness, confirming presence or dominance of specific invasive or protected species, calculating floristic quality, etc- may dictate the actual amount of sampling to be conducted in a survey. This is discussed at length in White Papers 1D, 1E, and 1F.**

Since this analysis did NOT clearly define the optimal number of survey sites for determining species richness, it was NOT applied to a comparable evaluation of the optimal number of sites for determining mean C values. An analysis for the minimal number of survey sites (rather than the optimal number of survey sites) is conducted at length for species richness in White Paper 1D and for coefficients of conservatism in White Paper 1F.

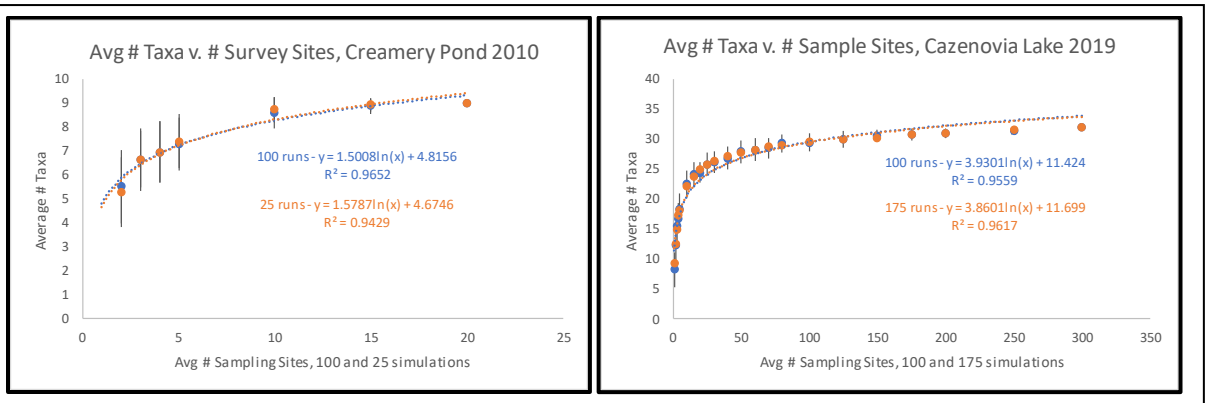
Section 2.5.3.3- Optimal number of computational runs for subsampling analysis

In addition to evaluating the optimal (most efficient) number of survey sites, the optimal number of survey runs (distinct combinations of randomly-chosen survey sites) can be evaluated through this ANOVA process by determining the number of runs required to stabilize the optimal number of sites. The optimal number of survey runs can vary significantly from lake to lake and annually within lakes, as seen in Table 2.5.3.3 for the PIRTRAM lakes with granular survey data. These data suggest that up to 100 runs (combinations of survey sites) should be sufficient to stabilize cumulative mean species richness estimates in all combinations of survey sites in smaller lakes (those with less than about 100 plant survey sites), while 100-200 runs are sufficient in larger lakes (those with 100 survey sites to about 300 survey sites).

It is not known if lakes with even more survey sites would require more than 200 runs to identify the optimized number of survey sites, since the raw granular data for these very large lakes are not available for this analysis. However, the expected distribution of species richness relative to survey sites does not appear to be strongly influenced by the number of sampling runs beyond 100 runs, even in larger lakes. These data further suggest both that a universal number of 100 runs can be used to evaluate the maximum number of taxa found in a lake, and that the actual number of survey runs does not significantly change the relationship between survey sites and number of taxa found in these surveys. This is apparent in Figures 2.5.3.1 and 2.5.3.2, which show the relationship between species richness and the number of survey sites based on varying number of simulations (“runs”) at each interval of survey sites for two lakes- Creamery Pond in 2010, and Cazenovia Lake in 2019. For Creamery Pond, 25 runs correspond to the minimum

number of runs required to stabilize the optimal number of survey sites; for Cazenovia Lake, 175 runs were required.

However, these figures show that the relationship between the number of recorded taxa and the number of survey sites is NOT dependent upon the number of computational runs used to generate these averages. The very small differences in these relationships between 25 and 100 runs (Creamery Pond) and 100 and 175 runs (Cazenovia Lake) are largely manifested in larger error bars (standard deviations) in the data associated with the smaller runs. The shapes of the regression plots associated with these runs do not change with the variation in the number of runs. Although these evaluations were not conducted on all PIRTRAM, the data in Figures



Figures 2.5.3.1 and 2.5.3.2- Relationship between Species Richness and Optimal # of Sampling Sites

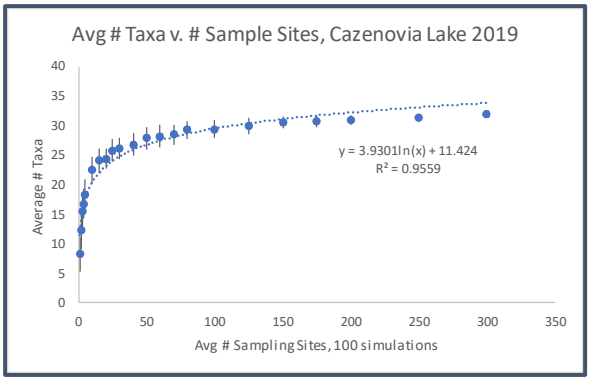
2.5.3.1 and 2.5.3.2 suggest that 100 runs should be sufficient to adequately characterize the relationship between observed taxa and number of sampling sites while minimizing the variance between these survey sites groups.

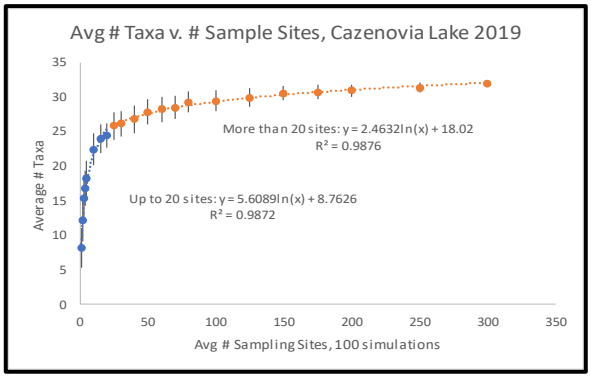
Therefore, 100 runs (also called "simulations") can be used to evaluate expected species richness and ultimately to calculate projected species richness in all lakes with granular plant survey data.

It is assumed that a comparable analysis of the number of computation runs needed to accurately project mean C values would similarly point to a 100 run analysis, but this was not conducted as part of this study.

Section 2.6- Regressions between cumulative species richness (or mean C values) and survey sites

Using the modified bootstrap analysis methods outlined above, cumulative mean (and standard deviation) values of species richness can be estimated for various intervals of survey sites, based on an analysis of existing granular survey site data in PIRTRAM lakes. This relationship- in most cases a logarithmic function showing an asymptotically increasing species richness as





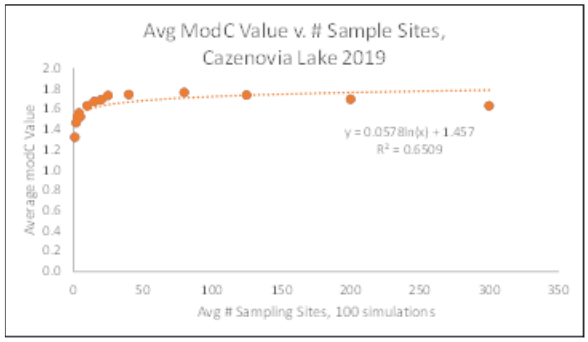
Figures 2.6.1 and 2.6.2- Example of species richness and single and split logarithmic regressions for Cazenovia Lake 2019

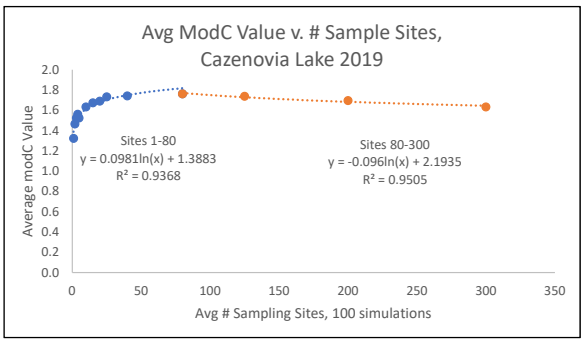
sample sites increasing- can be used to estimate species richness for both sub-optimal numbers of survey sites, and more importantly for a standardized larger number of aquatic plant survey sites, such as the 4 site per littoral acre standard proposed above.

For some lakes, a simple logarithmic relationship using each of the sampling site intervals (1, 2, 3, 4, 5, 10, 15….sites) described above appears to be sufficient to characterize the relationship between expected species richness and the average number of sampling sites. For these lakes, constituting the majority of the 55 PIRTRAM lake-years, an extrapolation of the logarithmic relationship to the projected number of survey sites recommended to optimize species richness calculations (= 4 sites per littoral acre) can be used to project species richness. For other lakes, a simple logarithmic relationship does not adequately describe this relationship, so a "split" regression is required. The point for splitting the regressions, or the number of plant survey sites separating the two regressions, can be graphically estimated as the inflection point of the descriptive curves. As noted above, this point corresponds to the optimal number of plant survey sites at which data concavity shifts, the point at which additional survey points result in a reduction in output per effort, and can be approximated through the ANOVA process described above. This is illustrated in Figures 2.6.1 and 2.6.2, using an example from Cazenovia Lake in 2019. In this example, it appears that a simple logarithmic equation describing the relationship between species richness and number of survey sites (Figure 2.6.1) does not adequately describe species richness at the higher end of the range of survey sites evaluated. A split logarithmic regression, shown in Figure 2.6.2, appears to more accurately project species richness at the 4 sites per littoral acre survey site frequency cited above.

For each of the PIRTRAM study lakes, single (Figure 2.6.1) or split (Figure 2.6.2) logarithmic regressions, comparing expected species richness and number of survey sites (using the mean of 100 survey runs), were generated using the modified bootstrap analysis to estimate species richness at varying numbers of survey sites. These regressions were extrapolated to the recommended maximum survey site density- 4 sites per littoral acre- to calculate a standardized projected species richness for each lake.

These regressions (either single or split) are outlined in White Paper 1D, Appendix 3.2.1 for each of the PIRTRAM lakes with granular plant survey data (Table 2.5.3.1). The regression equations shown in White Paper 1D, Appendix 3.2.1 can be used to estimate projected species richness (given a standardized high density of survey sites, = 4 sites per littoral acre, as discussed below) and provide estimates of percentage of projected species richness given various numbers of survey sites. These data are discussed at length in White Paper 1D- Species Richness.





Figures 2.6.3 and 2.6.4- Example of mean C value single and split logarithmic regressions for Cazenovia Lake 2019

As for coefficients of conservatism and mean C values, Figures 2.6.3 and 2.6.4 shows the change in single and split regressions of mean (modified) C values as the comparable number of survey sites increase. The differences in these regressions among PIRTRAM lakes is discussed at length in White Paper 1F, but Figures 2.6.3 and 2.6.4 also indicate that a standardized survey site density (and associated mean C_m values used in FQI calculations) will also improve comparison among lakes and over time in single lakes.

Section 2.7- Estimates of species richness (and mean C values) based on standardized projected survey sites

The logarithmic regressions described above can be used to estimate the species richness for any specific number of aquatic plant survey sites, and as noted above are provided in Table 2.5.2.1. The tables in White Paper 1D, Appendix 3.2.1 show the estimates of species richness for various intervals of aquatic plant survey sites- existing number of survey sites, and the number of survey sites corresponding to sites per littoral area, including 4 littoral hectares per site, 2 littoral hectares per site, 1 littoral hectare per site, 0.5 littoral hectares per site (or 2 sites per littoral hectare), 0.25 littoral hectares per site (or 4 sites per littoral hectare), and 0.25 littoral acres per

site (or 4 sites per littoral acre). The latter site density- 4 sites per littoral acre- corresponds to the recommended standardized survey site density outlined above, which also corresponds to the site density roughly achieved in the most intensive aquatic plant surveys conducted through PIRTRAM. This survey site density, referred to here as the "standardized projected survey site" density, allows for a comparison between lakes and across the PIRTRAM aquatic plant survey programs.

A perusal of the PIRTRAM lakes in White Paper 1D, Appendix 3.2.1 shows a high correlation between observed species richness (based on the actual number of aquatic plant survey sites) and predicted (projected) species richness based on the logarithmic regressions provided for each lake. This suggests that projected species richness values (based on extrapolation of these regressions) would likely be similar to observed species richness if the recommended standardized high density of survey sites (= 4 sites per littoral acre) were actually sampled. However, it is possible that these projections overestimate "maximum" species richness, since other factors (including plant competition in limited space or substrate, specific discrete numbers and types of unique taxa introduced to a lake ecosystem) may be the limiting factor in determining species richness.

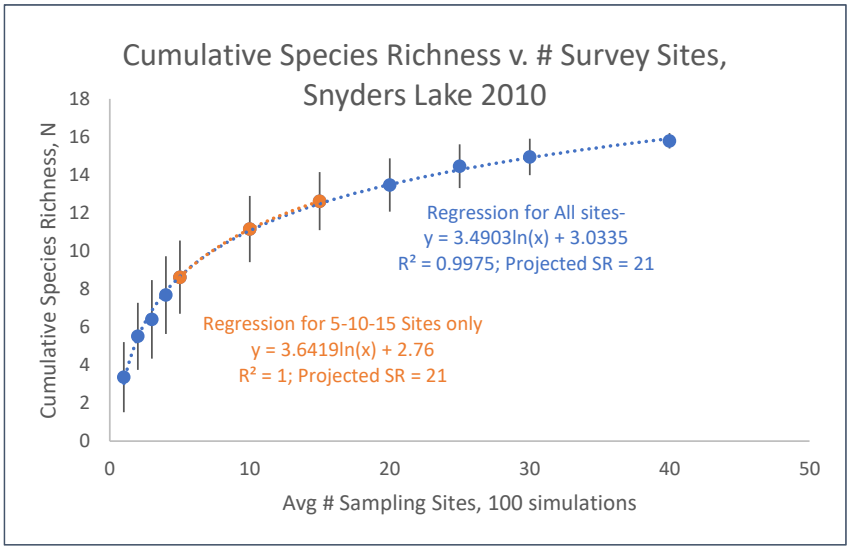
Nonetheless, the methods outlined in this White Paper and summarized in White Paper 1D provide a standardization of survey site densities, a process for estimating species richness at defined numbers of survey sites. This results in a maximum species richness estimate, allows for a comparison between programs, within programs, and on individual lakes from year to year. As seen in White Paper 1F, this projected species richness could also be applied to floristic quality indices (FQIs), but there is insufficient data in the PIRTRAM dataset to develop those enhanced FQI estimates. However, this should ultimately be a goal of future monitoring programs and future FQI evaluations.

The same method can be used to determine the mean C values at a standardized survey site density; the application of this method is discussed at length in White Paper 1F.

Section 2.8- Estimates of species richness based on reduced or minimal survey sites.

Projected species richness estimates are generated from data associated with a large number of plant survey sites displayed in White Paper 1D, Appendix 3.2.1 for the PIRTRAM lakes, an estimate of the maximum number of plant taxa can be generated from logarithmic regressions of cumulative species richness from truncated surveys using smaller numbers of plant survey sites. While ideally all aquatic plant monitoring programs would survey plants using a survey site density of 4 sites per littoral acre, as described above, this is probably unrealistic given available time and resources. Moreover, it is possible to estimate maximum species richness using even fewer survey sites than are used in nearly all of the PIRTRAM surveys.

Figure 2.8.1 -Example of Full and Truncated Species Richness Regressions

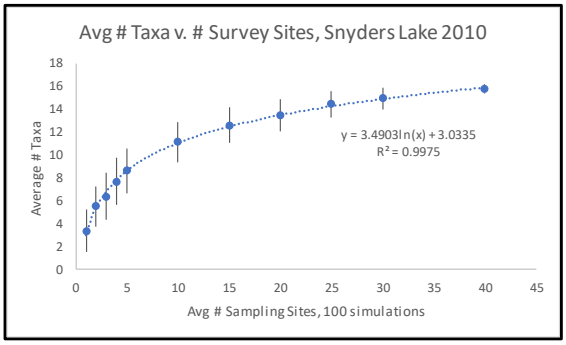


There is a great advantage in time and resources expended to use a smaller number of survey sites to estimate overall species richness for the lake, AS LONG AS the errors introduced in truncating the number of sampling sites are small overall (indicating relatively high accuracy and small errors relative to the gain in time and resources), allowing for truncated plant surveys to be included in routine monitoring programs.

One method for estimating the maximum number of unique plant taxa in a lake is to extrapolate the logarithmic regression of the cumulative number of unique taxa in smaller discrete numbers of plant survey sites to the number of survey sites associated with a 4 site per littoral acre grid. So, for example, a regression of the expected number of unique taxa in 1, 2 and 3 sites in Ballston Lake (or any other three consecutive number of survey sites) can be extrapolated to estimate the number of plant taxa in 474 plant survey sites (= 4 sites per acre of littoral area) and compared to the calculated estimate of plant taxa using the regression of the entire number of plant survey sites (as shown in White Paper 1D, Appendix 3.2.1). This allows for an estimate of the amount of sampling effort required to estimate overall (maximum) species richness. While the estimates of species richness with these small numbers of survey sites are presumably not as accurate as the estimates from, in the case of Ballston Lake, 30 survey sites, this analysis can provide information about the error in this truncated estimate relative to the much larger sampling effort required to gain a more accurate estimate of overall species richness. The same process can then be conducted stepwise in increasing numbers of survey sites until the error (mean number of unique plant taxa and variance) is sufficiently small.

Figure 2.8.2- Example Species Richness and Survey Effort Calculations

Snyders Lake 2010 Plant Survey conducted by DEC DOW Albany



%Max Taxa	#Sites
10	1
20	1
30	2
40	4
50	7
60	13
70	23
80	42
90	76

	0.25 ac per site	0.25 ha per site	0.5 ha per site	1 ha per site	2 ha per site	4 ha per site	Existing Survey	Modeled Existing
# Sites (X)	148	60	30	15	7.5	3.75	44	44
Taxa in X sites	20.5	17.3	14.9	12.5	10.1	7.6	16	16.2
% Max	100%	85%	73%	61%	49%	37%	78%	79%

(Maximum number of sites and taxa correspond to site frequency of 0.25 acres per site)

An example of this process is provided in Figure 2.8.1 for Snyders Lake, which shows the distribution of cumulative species richness as a function of the survey sites deployed at the lake in 2010. These data show an observed species richness of 16 unique taxa found in 40 actual survey sites. Based on the logarithmic regression of all the aquatic plant survey sites, a projected species richness of 21 (actually 20.5) unique taxa can be calculated based on a standardized survey site density of 4 sites per littoral acre, corresponding to 148 survey sites. However, when only 15 survey sites (uniformly distributed across the littoral zone) are used, a "truncated" logarithmic regression of the expected cumulative species richness at 5 sites, 10 sites, and 15 sites also calculates a projected species richness of 21 (21.0) unique taxa when this regression is extrapolated to a survey site density of 4 sites per littoral acre (=148 survey sites). The same process can be used to estimate species richness at various intervals of survey sites up to the standardized 4 sites/littoral acre. An example of this is provided in Figure 2.8.2, also for Snyders Lake, and all of the PIRTRAM lakes regressions and expected species richness values are summarized in White Paper 1D, Appendix 3.2.1. The analysis of these data is provided in White Papers 1D-Species Richness, 1E-Evaluation of Individual Species, and 1F-Coefficients of Conservatism.

The same tools can also be used for evaluating and calculating coefficient of conservatism. The application of the tools outlined in this section for generating C values are discussed at length in White Paper 1F.

Section 3- Tools Used in the Evaluation of Coefficients of Conservatism (White Paper 1F) only….

Section 3.1- Background

Section 2 summarizes the tools and methods used in evaluating species richness and, in some cases, coefficients of conservatism, including subsampling and bootstrapping methods to develop cumulative species richness and mean C value values projected to any survey site density, including the proposed standardized survey site density of 1 site per littoral hectare. The application of these methods is detailed in White Paper 1D and White Paper 1F as applied to those PIRTRAM lakes with granular survey site data, with a focus on mean C values modified for a proposed modified C value system (mean C_m values).

However, as discussed at length in White Paper 1F, coefficients of conservatism can and should be modified using an alternative C value scale, and should also be corrected for plant frequency and/or plant abundance. The tools and methods used for these modifications and corrections are described below, and in more detail in White Paper 1F.

Section 3.2- Modifications to the C Value System and Mean C Values

The most frequently used measure of floristic quality is the floristic quality index, as shown in Equation 3.2.1 below (reproduced from Equation 1.1 in White Paper 1F):

Equation 3.2.1: $FQI = \bar{C} \times \sqrt{N}$, *and* $\bar{C} = \Sigma C / N$; where
N = number of unique plant species in a lake (=observed species richness, or oSR), and
C = coefficient of conservatism for each unique species (= C value)

The mean C ($\bar{C}$) value used in this equation is typically derived from C values assigned to all native plants using a 10 point scale, ranging from the most ecologically tolerant plants (C = 1) to the least ecologically tolerant plants (C = 10). All exotic species, regardless of their invasiveness, are assigned a C value of 0. C values have been assigned to terrestrial and aquatic plants in many states, including New York, through an extensive process overseen by the state Natural Heritage Programs. While the assigned values in New York (or C_{ny} values) allow for a differentiation in ecological quality among nearly all plants found in aquatic plant surveys, these assignments do not always align closely with the results from these surveys (in part because the system was largely developed for and applied to terrestrial plant community assessments).

The issues with the New York C value system as it relates to New York lake aquatic plant surveys are discussed at length in White Paper 1F, but can be summarized as follows:

- Plant survey results and C value assignments can be strongly affected by sampler experience and expertise, dissuading the use of many surveys (including volunteer-based surveys) in floristic quality assessments
- Aquatic plant identification training, particularly to support floristic quality assessments, require gained expertise in as many as 1200 unique aquatic plant species

- High plasticity in some aquatic plants increases uncertainty in accurate aquatic plant identifications
- Plants collected in aquatic plant surveys require collecting and retrieving deep submergent aquatic plants, resulting in tentative identifications of plants without all defining characteristics (thereby increasing identification errors)
- The time of year for surveys may influence the ability to collect complete plant reproductive structures and other plant parts (flowers, emergent leaves, turions, subterranean tubers, etc.), but must be optimized to collect most plants
- Plants collected in some historical datasets are not identified to species level, or present programs identify plants in some habitats (submergent macrophytes) to species level, and other habitats (floating and emergent macrophytes, and submergent macroalga) to genera, precluding the assignment of C values to some collected and reported plants
- Invasive species are assigned a 0 value, regardless of invasiveness
- The C value and resulting FQI scale may be difficult to interpret, particularly for aquatic plant communities
- Coefficients of conservatism are assigned the same value regardless of the frequency or abundance of these plants

Many of these issues can be addressed by creating an alternative C value system that addresses plant collection and identification issues (by reducing the number of plant species or genera requiring a very high that a plant is NOT one of a few plants rather than a plant is a specific plant), issues relating to differing habitats subject to species- (versus genera-) level identifications (by assigning nearly all native benign species and genera to a single C value), issues related to relative invasiveness (by assigning different C values based on invasiveness) and challenges in interpreting data (by assigning clear and distinct delineations, rather than gradations, between good and bad plants).

White Paper 1F outlines a process by which a modified C value system (or C_m) is deployed, assigning all plants to a C value scale that ranges from -5 to +5. The

-5 = very highly invasive (non-native) plants
-3 = moderate to highly invasive (non-native) plants, including regionally invasive plants
-1 = non-native plants with low invasiveness
+1 = nuisance native plants
+3 = benign (beneficial) native plants
+5 = protected (rare, threatened, or endangered) native plants

The modified C_m value scale exploits two regulatory lists adopted in New York State. The Protected Plant List (rare, threatened, endangered, and exploitably vulnerable species), reported in 6 NYCRR 193.3 (https://www.dec.ny.gov/docs/wildlife_pdf/2019rareplantlists.pdf) identifies those high value aquatic plants that warrant protection. In addition, the Regulatory System for Non-Native Species (http://www.dec.ny.gov/animals/63402.html) characterizes the invasiveness of all non-native plants in New York state for the purpose of establishing restrictions on the sale, transport, and possession of these plants.

White Paper 1F provides more details about the modified C_m value system, including rationale for the proposed use of this modified system, ranges of mean C_m values associated with surveyed lakes, changes in these values from year to year and in response to management, and a proposed scoring system for evaluating floristic quality based on mean C_m values. In addition, Appendix 2.1 in White Paper 1F offers suggested C_m values for each assigned C_{ny} value and associated aquatic plant species or genera.

Section 3.3- Frequency Corrections to Mean C Values

As noted above, traditional measures of floristic quality- specifically, the coefficients of conservatism- do not account for plant frequency or abundance. The resulting FQI calculations do not consider the number and relative abundance of plants (as opposed to the number of plant species), likely resulting in differences in actual floristic quality- ecosystem function, sediment retention, fish habitat, recreational impediments, etc- despite similarities in calculated floristic quality.

The weighting factors associated with plant frequency (and abundance, summarized below) can be assigned to the C_m values in the FQI equations provided in Equation 3.2.1, since the weighting would influence the quality of the plant community rather than the number of plant species. These weighting factors can be used to evaluate mean C_m values corrected for relative or absolute frequency:

a. *Relative or normalized frequency* refers to a means for evaluating those plants that occur at a higher frequency than other plants, regardless of the absolute frequency. The formula used to calculate normalized weighted frequency mean C_m values is as follows:

   Equation 3.3.1: C_{m_nf} = *sum of (all sites counts x* C_m *value for species) / sum of all sites taxa counts*

   where "m" refers to modified, "n" refers to normalized and "f" refers to frequency

b. *Absolute or unbounded frequency* refers to the means for evaluating those plants that are more frequently found than other plants, regardless of the relative frequency. These corrections can be calculated by taking the sum of all species counts x the C_m value for each species (the numerator in Equation 3.3.1), and divide this by the "opportunities" for plant frequency, resulting in Equation 3.3.2:

   Equation 3.3.2: C_{m_uf} = *sum of (all sites counts x* C_m *value for species) / (number of plant species x number of survey sites*).

   where "u" refers to unbounded frequency

Either equation can be easily applied to the entirety of survey site results using observed C values. **However, absolute or unbounded frequency corrections are much more easily applied to projected individual (component) and mean (community) C_m values, as discussed in White Paper 1F, so modified C values corrected for unbounded frequency**

(C_{m_uf}) are calculated for the PIRTRAM dataset. This information is reproduced in White Paper 1F.

Section 3.4- Abundance Corrections to Mean C values

As noted above, the standard formula for determining projected mean C_m and FQI values does not account for plant abundance. As with plant frequency, plant abundance can be evaluated as a relative (normalized) or absolute (unbounded) factor. The weighting factors associated with plant abundance should also be assigned to the "component" (individual plant) and mean (overall aquatic plant community) C_m values, as with frequency-based factors, since the weighting would influence the quality of the plant community rather than the number of plant species.

Plant abundance was estimated at nearly all of the lakes surveyed in the PIRTRAM aquatic plant dataset, using the previously cited US Army Corps of Engineers and Cornell/SUNY Oneonta relative abundance scales, applied to two-sided rake toss data. These relative abundance assessments culminated in a summary of relative abundance scales in Table 3.4.

Table 3.4: Plant Abundance Categories Used in NYS Plant Surveys

Density Category	Estimated Quantity from Average of 1-2 Rake Tosses	Approximate Biomass	Assigned Score
No plants (Z)	Nothing	0 g/m^2	0
Trace (T)	Fingerful (of plants)	up to 0.1 g/m^2	1
Sparse (S)	Handful	0.1 to 20 g/m^2	5
Medium (M)	Rakeful	20 to 100 g/m^2	25
Dense (D)	Can't Bring In Boat	100 to 400 g/m^2	125

Reference: Kishbaugh, 2020; Johnson, 2008

The approximate biomass associated with each density category was generated from multiple paired rake toss and quadrant biomass sampling conducted at Chautauqua Lake (Johnson, 2008). The assigned score in Table 3.4 represents a $\log_5$ scale representing the relationship between a density category and approximate biomass (Kishbaugh, 2020). Other researchers may elect to choose a different scale for defining the weighted distinction between density categories used in the PIRTRAM method and in Table 3.4, but it is not anticipated that the results discussed below would change significantly in response to using this alternative weighting scale.

As with plant frequency corrections, relative abundance measures can be used to corrected component and mean C_m values, as summarized below and in more detail in White Paper 1F:

a. *Relative or normalized abundance* refers to a means for evaluating those plants that occur at a higher abundance than other plants, regardless of the absolute abundance. The formula used to calculate normalized weighted abundance mean C_m values is as follows:

Equation 3.4.1: C_{m_na} *= sum of (all sites abundance x* C_m *value for species) / sum of all sites taxa abundance*

where "a" refers to abundance (the other terms are defined in Equation 3.3.1)

b. *Absolute or unbounded abundance* refers to the means for evaluating those plants that are more abundant than other plants, regardless of the relative abundance. The same general method used for evaluating absolute plant frequency is also applied here for evaluating absolute plant abundance, and is described in Equation 3.4.2:

Equation 3.4.2: C_{m_us} = *sum of (all sites abundance x* C_m *value for species) / (number of plant species x number of taxa)*

As with frequency corrections**, absolute or unbounded abundance corrections are much more easily applied to projected individual (component) and mean (community) C_m values, as discussed in White Paper 1F, so modified C values corrected for unbounded abundance (C_{m_ua}) are calculated for the PIRTRAM dataset.**

References:

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. Monograph No. 38, Conference Board of the Mathematical Sciences - National Science Foundation regional conference series in applied mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics; 92 p.

Kishbaugh, S.A. 2022. White Paper 1A- New York state plant survey programs. Unpublished white paper

Kishbaugh, S.A. 2022. White Paper 1B- Elements of plant surveys. Unpublished white paper.

Kishbaugh, S.A. 2022. White Paper 1D- Species richness in New York state lakes. Unpublished white paper.

Kishbaugh, S.A. 2022. White Paper 1E- Plant lists and individual species in New York state lakes. Unpublished white paper.

Kishbaugh, S.A. 2022. White Paper 1F- Coefficients of conservatism in New York state lakes. Unpublished white paper.

Kishbaugh, S.A. 2022. White Paper 1G- Floristic Quality Indices in New York state lakes. Unpublished white paper.

Smith, W.P.; D.J. Twedt,; R.J. Cooper,; D.A.Wiedenfeld; P.B. Hamel and R.P.Ford, 1995. Sample Size and Allocation of Effort in Point Count Sampling of Birds in Bottomland Hardwood Forests. In: Ralph, C. John; Sauer, John R.; Droege, Sam, technical editors. 1995. Monitoring bird populations by point counts. Gen. Tech. Rep. PSW-GTR-149. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station: p. 7-1.