Summary Report of Version 1 and Version 2 Data and Findings from Assessment Protocol Implementation

Deliverable 2.5, Raingarden and Bioretention Assessment Protocol

Attachments: Appendices 1-8

Introduction:

This document represents a summary of data collected in versions 1 and 2 of the Bioretention/Rain Garden Facility Assessment Protocol ("the protocol") Implementation, an analysis of the resulting data, and recommendations for the final assessment protocol. In order to provide a comprehensive project data summary in one document, portions of this report duplicate information found in Deliverable 2.3 Technical Memo: Results from Version 2 Assessment Protocol Implementation.[2]

At the outset of the project a review of rain garden and bioretention facility studies and assessment strategies in the published literature was conducted. It focused on three main goals: 1) identify rain garden and bioretention functions, their field indicators and potential field- based monitoring procedures; 2) review industry-standard designs, construction and installation practices, and maintenance activities that influence their function, and; 3) review survey methodologies to gather information on landowner attitudes and perspectives on rain garden and bioretention function and acceptance.[1] The intentions of the project are to: (1) develop volunteer / staff-friendly data collection methods that can be implemented across western Washington, that do not need extensive equipment or access to lab facilities (2) collect defensible data, regardless of who collects it; and (3) provide an initial assessment of rain gardens and bioretention function and acceptance. ; (4) better understand landowner values about rain gardens and rain garden maintenance incentives through a social science survey; [1]

Methods Overview:

The priorities for the data collection aspects of this project were ease of implementation, repeatability across geography, data consistency across different implementers and overall data quality and utility in terms of scientific and facility management value. Feedback from implementers was collected after the first and second protocol versions were implemented, including the full assessment and a rapid assessment version of the second protocol. Feedback was collected via direct communication with implementers. Between version 1 and 2, approximately 20% of the metrics were removed from version 1's relatively exhaustive list of metrics. For the rapid assessment protocol another 38% of the metrics were removed including particularly time-intensive metrics like determining ponding volume of the rain garden or bioretention facility.[2]

A total of 35 volunteers in three counties (Snohomish, Thurston and Jefferson) received approximately 8 hours of training on protocol 1. Implementers worked in teams of 2-3 individuals and assessed 14 sites, with each site repeated by a different team of volunteers to assess repeatability. For Version 2, 77 volunteers in four counties (Pierce, Snohomish, Thurston and Jefferson) received 8 hours of training. An additional 6 volunteers were recruited to implement assessments without any formal training. These "untrained" volunteers were given the identical instructions that were developed with the assessment protocol and used by trained volunteers. Volunteers, working in teams of 2-3 assessed 41 sites, with most sites repeated with either the rapid assessment or full assessment by a different team of volunteers to assess repeatability. The breakdown of the different assessments are as follows and in Table 1 and Figure 1.[2]

- 35 Rapid and 47 Full assessments were completed in all counties.
- 67 assessments were completed by trained volunteers and 15 completed by untrained volunteers.

Table 1: Summary of assessment types and training of volunteers using version 2 protocol.[2]

Training Level	Full Protocol (47)	Rapid Protocol (35)
Trained (67)	47	20
Untrained (15)	0	15

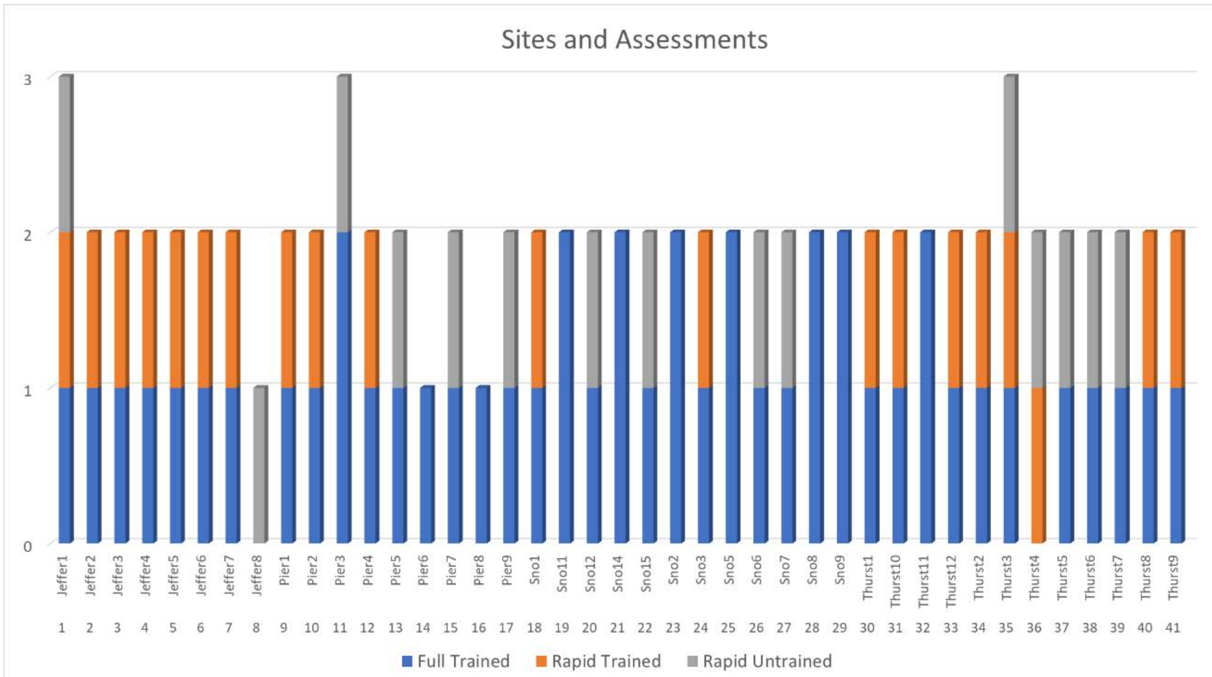


Figure 1: All sites monitored with type of assessment (Full vs Rapid) AND level of training (Trained vs Untrained) using version 2.[2]

Version 1 of the assessment protocol had an approximate total of 212 parameters (this figure varied depending on the # of different plant species in the facility) (Appendix 1), which was narrowed to 170 parameters in version 2 (Appendix 2) and 106 for the rapid assessment (Appendix 3). The refinements were based project goals, feedback from implementers, utility of the data for assessment purposes, ability to assure accuracy, and recommendations from the project technical advisory committee (TAC) [3]. The original data for versions 1 and 2 are presented in appendices 4 and 5, respectively.

The rapid assessment protocol was created by removing some of the metrics (and the related instructions for those metrics) from version 2 of the full protocol, each of the metrics in the rapid protocol has an identical analogue in the full protocol. The rapid assessment protocol contained identical instructions and data collection reporting as the full protocol for all the metrics that it included. To assess the replicability of methods and comparing and contrasting across training and protocols, we conducted analyses comparing overall similarity of different assessments at the same site (using an analysis method called "cosine similarity") and overall variation and distribution of data for each metric across all sites assessed.[2]

2. Assessment protocol implementation challenges and recommendations

- Determining ponding volume using simple field equipment proved to be difficult for many of the implementers at many of the sites. This information was also not considered valuable if the contributing area or a measured infiltration rate were not available. This parameter should be eliminated for widespread use of this assessment protocol.
- Assessing soil compaction using simple field equipment did not provide consistent results between implementers. Potentially due to variability in different penetrometers or user errors, as well as differences in soil moisture content on different assessment dates. A coarser measurement strategy is recommended.
- Data collection via hand written forms that are subsequently transcribed into a database can provide errors if not done carefully in a consistent manner. The collection database should provide drop down menus for measurements to eliminate some sources of inconsistencies. If possible, a mobile device-based data entry tool (mobile app) that mirrors the hard copy forms, includes drop down menus for each field, and links automatically to a database would be the optimal solution given current technology, and would remove the undesirable and error-prone step of data transcription.
- Management of photo point information was challenging and time consuming in version 1. Our photo management and tracking form is included as an appendix that we recommend as optional at this time. It may be useful to a user who has a strong need and may help determine is a field visit is necessary. Photographic mobile device technology is continuing to improve, yet varies in its accessibility for different implementers, however photographs could be well-suited to integrate into mobile app version of this protocol.

3. Replicability of results across different implementers and dates/times

Site-by-site

Using cosine similarity (a statistical measurement of similarity across several variables) to compare how similar assessments of the same site were when made by different teams and looking across all metrics combined, our team found that the site assessments were generally comparable. Cosine similarity across the repeated sites (summarized in Table 2) were consistently skewed towards 1 (100% similarity). In contrast, cosine similarity values of different sites showed relatively normal distributions of values as expected when comparing non-identical sites and indicates how much variability the protocol is capable of measuring. These cosine similarity results support the assertion that the assessment protocol in both forms is highly repeatable and captures consistent quantitative data on rain gardens and bioretention facilities. These results also support the assertion that the assessment protocol can provide an overall indication of a specific rain garden or bioretention facility's current state. The degree of replicability appears consistent within and between the different protocol versions (full and rapid) as well as between trained and untrained volunteers applying the rapid protocol. We did not account for time between assessments and some assessments were conducted a few weeks apart from each other. Therefore, actual changes in the condition of individual sites over time is a potential factor that would lower cosine similarity scores despite good repeatability of the protocol.[2]

Table 2: Summary of Cosine Similarity Values for repeated sites and different sites

Type of repeated assessment	Median of repeated site similarity (range) and sample size	Median of different site similarity (range) and sample size
Full Protocol, Trained Volunteers replicates	**0.78** (0.49-0.97) n=8	**0.5** (0.14-0.92) n=56
Full Protocol + Trained Volunteers vs. Rapid Protocol + Trained Volunteers	**0.78** (0.32-0.95) n=18	**0.70** (0.15-0.96) n=306
Full Protocol + Trained Volunteers vs. Rapid Protocol + Untrained Volunteers	**0.81** (-0.05-0.95) n=12	**0.52** (-0.008-0.88) n=132
Rapid Protocol + Trained Volunteers vs. Rapid Protocol + Untrained Volunteers	**0.94** (0.84-0.97) n=3	**0.675** (0.58-0.9) n=6

In addition, we used a proportioned difference analysis of the assessments of 8 sites to determine variability between different fully trained assessors using the version 2 protocol on the same site. In this analysis, displayed in figures 2 and 3, the closer the median is to 1.0, the less similar the reporting of that metric. So for example, in the hydrology figure, cover of big rock (h_cover_bigrock) and pea gravel (h_cover_peagravel) were reported very similarly, while mulch cover (h_cover_mulch) and depth to native soils (h_native_depth_z1) show greater variability.

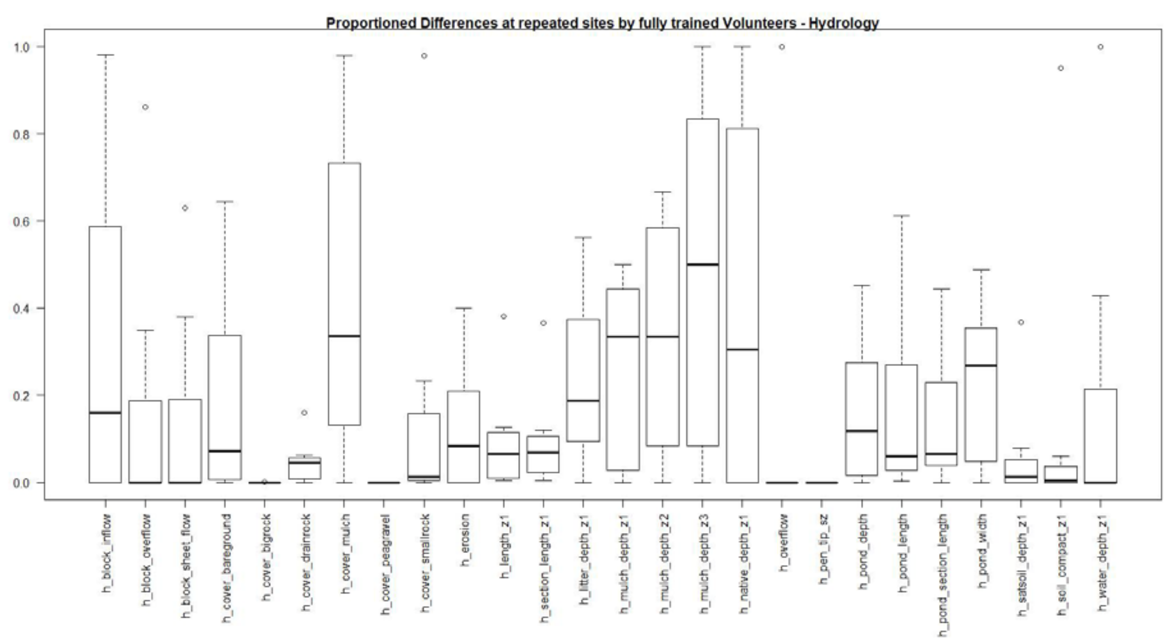


Figure 2: Proportioned differences of hydrology parameters at repeated sites by fully trained volunteers.

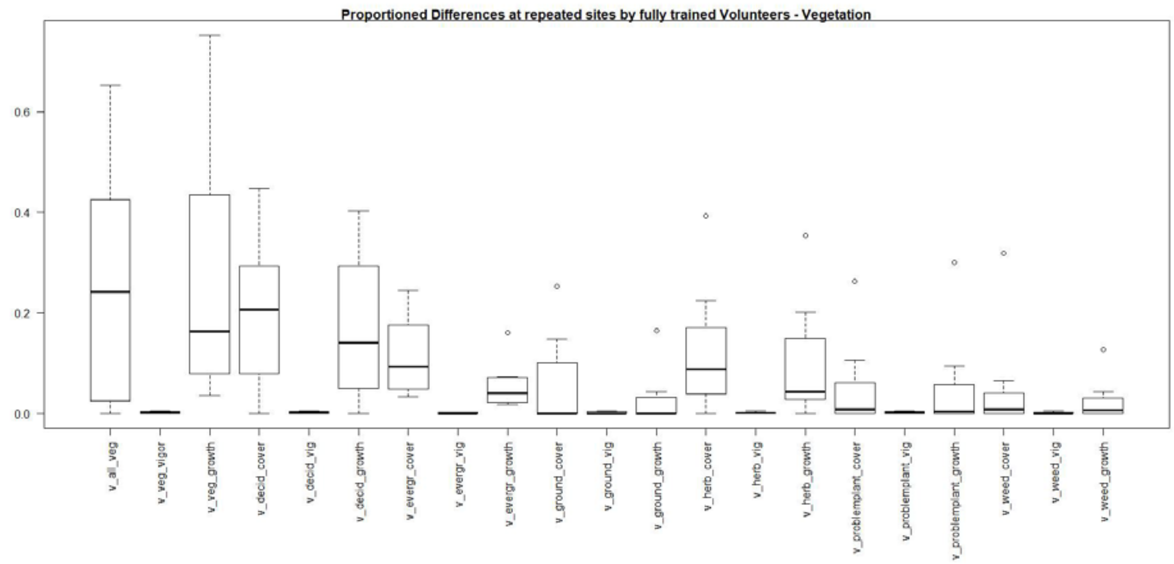


Figure 3: Proportioned differences of vegetation parameters at repeated sites by fully trained volunteers.

The project team decided not to eliminate any of the parameters in version 2 as a result of this analysis, however some parameters and their measurement techniques were modified or eliminated for other reasons, including technical advisory committee recommendations and the input of assessors. The recommended changes to version 2 of the protocol are outlined in appendix 8.

Metric-by-Metric Assessment

In order to assess how informative each metric is we looked at the distributions of values for each metric pooled for all sites and all assessments using a data transformation that scaled each metric to a -100 to +100 range according to that metric's ability to inform the overall effectiveness of a facility. This transformation was based on the collective expertise of our project team and effectively calibrated each metric such that we can see which metrics provided the scores that inform us the most and least in terms of positive function, e.g. presence of an overflow ("h overflow" in figure 4) which was deemed highly important and was consistently scored as present had a median score (horizontal line) of +100 (for overflow present), but also received 0 and -100 scores for unknown and absent respectively in one or more cases). This analysis was conducted once for metrics that are considered important for hydrologic function (Figure 4) and again for metrics considered important for vegetation (and related to plant community, plant health, ecological function, and habitat function) (Figure 5).[2]

The majority of the metrics relating to the hydrological function of rain gardens and bioretention facilities averaged close to zero on our transformed scale and showed relatively narrow ranges of variability. This distribution of data is not overly surprising for two reasons. First because there are very few metrics that can be measured on any given day (i.e. not requiring observation/measurement after a specific weather event) that will tell you if a rain garden or bioretention facility is hydrologically failing or thriving (our literature review indicated that such was the case). Secondly, relatively few rain gardens and bioretention facilities, even those where the vegetation is in poor condition, fail hydrologically, a point that was also generally supported in our literature review. This analysis will be used in determining which metrics may be removed in the final (3rd) version of the protocol, as well as which are most important to retain for the rapid version of the protocol.[2]

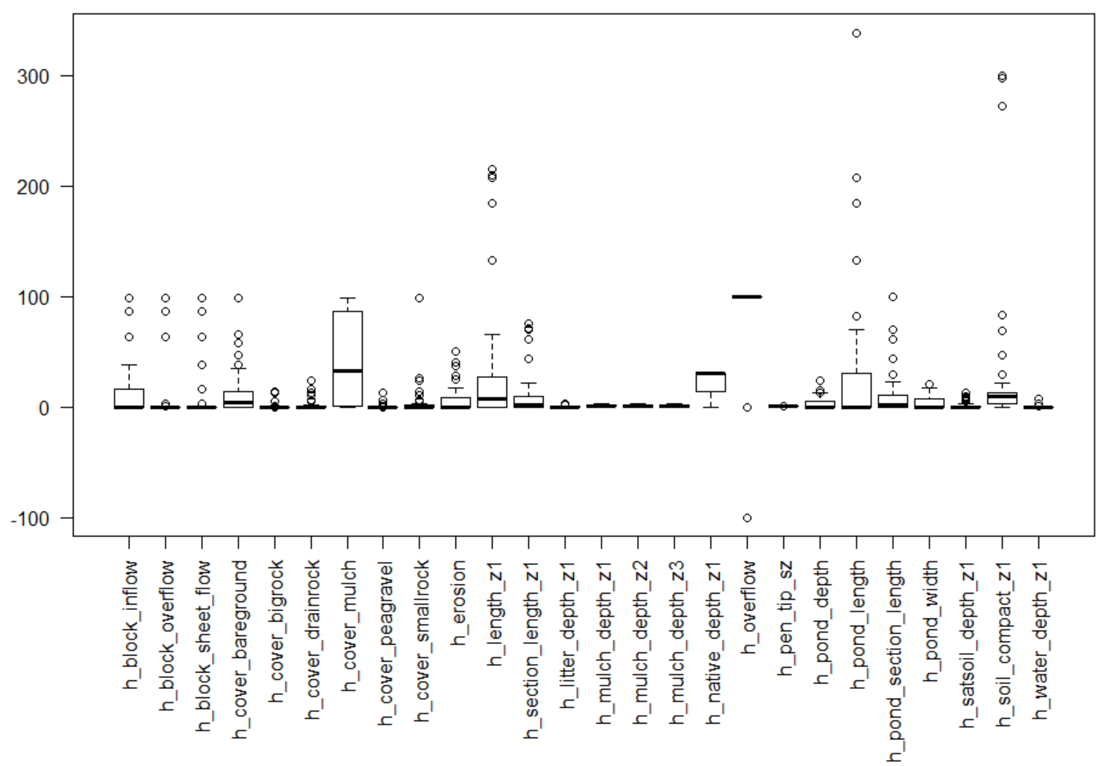


Figure 4: Boxplot showing the variability of all hydrology-related metrics' values at all sites aggregated across all zones of the bioretention cell/rain garden. Very few of the metrics produced normally distributed of data (i.e. bell curves) and the vast majority of data points for most metrics were at or near zero (i.e. Poisson distribution).[2]

Metrics related to vegetation functions scored relatively higher than hydrology-related metrics and also showed greater overall variability and more normal distributions (Figure 5). No one class of vegetation (evergreen, deciduous, ground cover, herbaceous, weeds and "problem plant" etc.) provided more overall information about vegetative functions of rain gardens or bioretention facilities. Cover and growth metrics had stronger correlations to vegetative function than did metrics of vigor according to our analysis. This analysis will further inform the metrics selected for version 3 of the protocol and rapid protocol.[2]

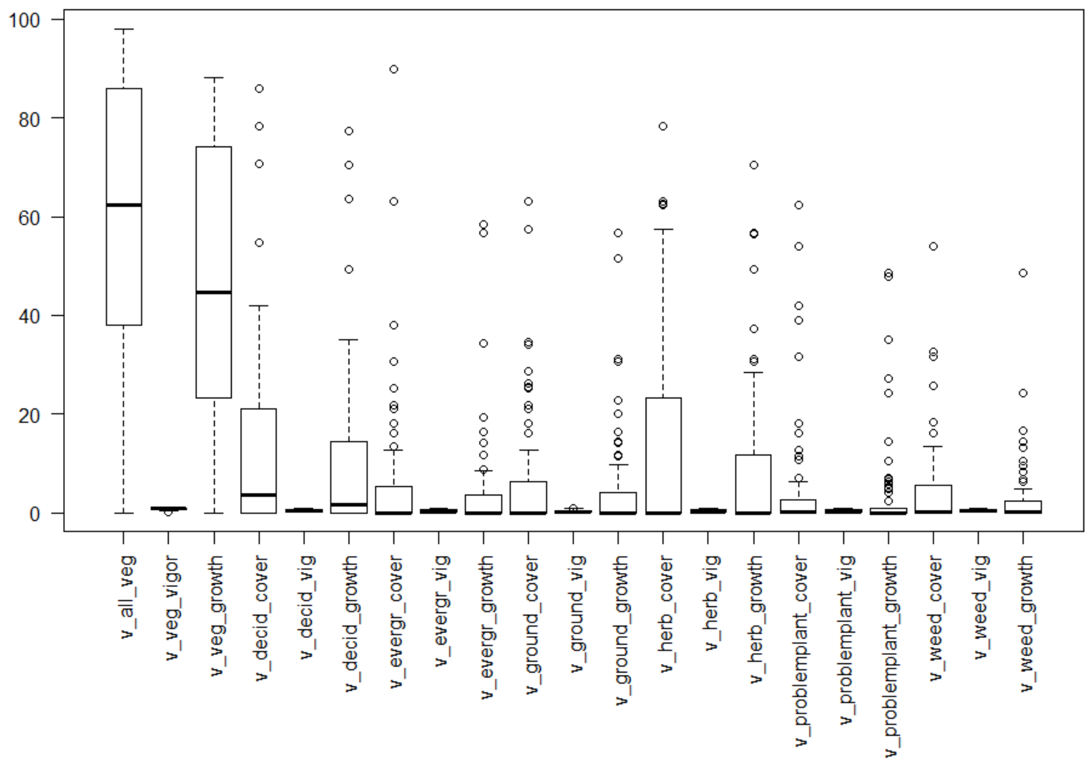


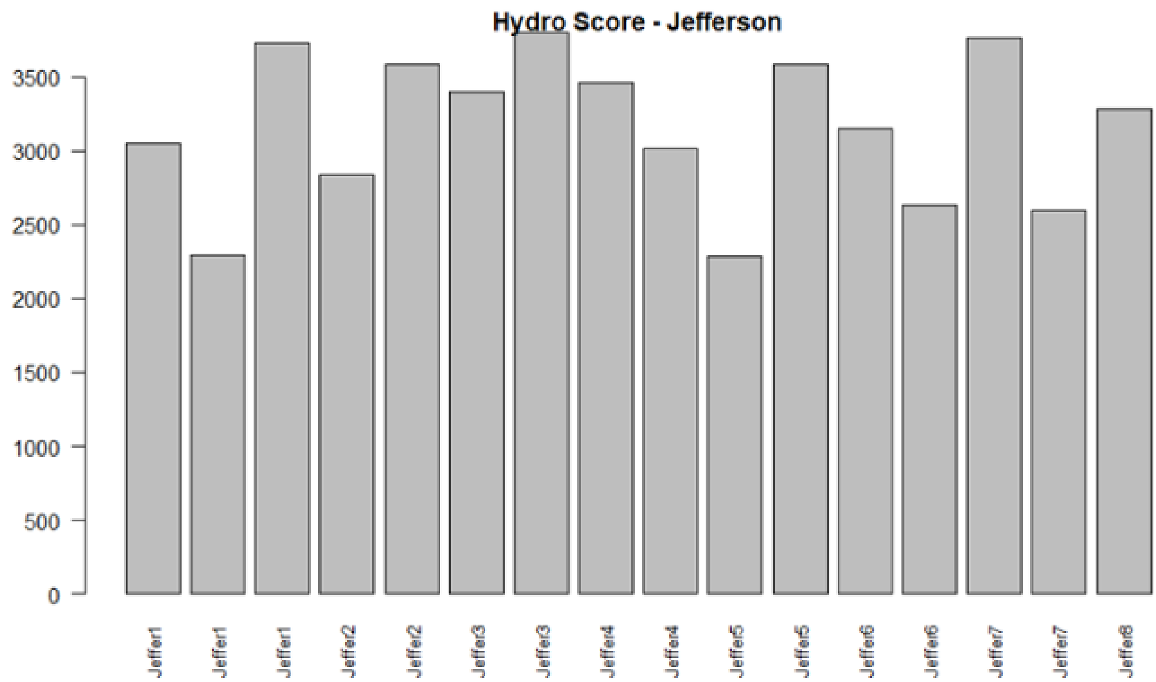
Figure 5: Boxplot showing the variability of all vegetation scores at all sites aggregated across all zones of the bioretention cell/rain garden.[2]

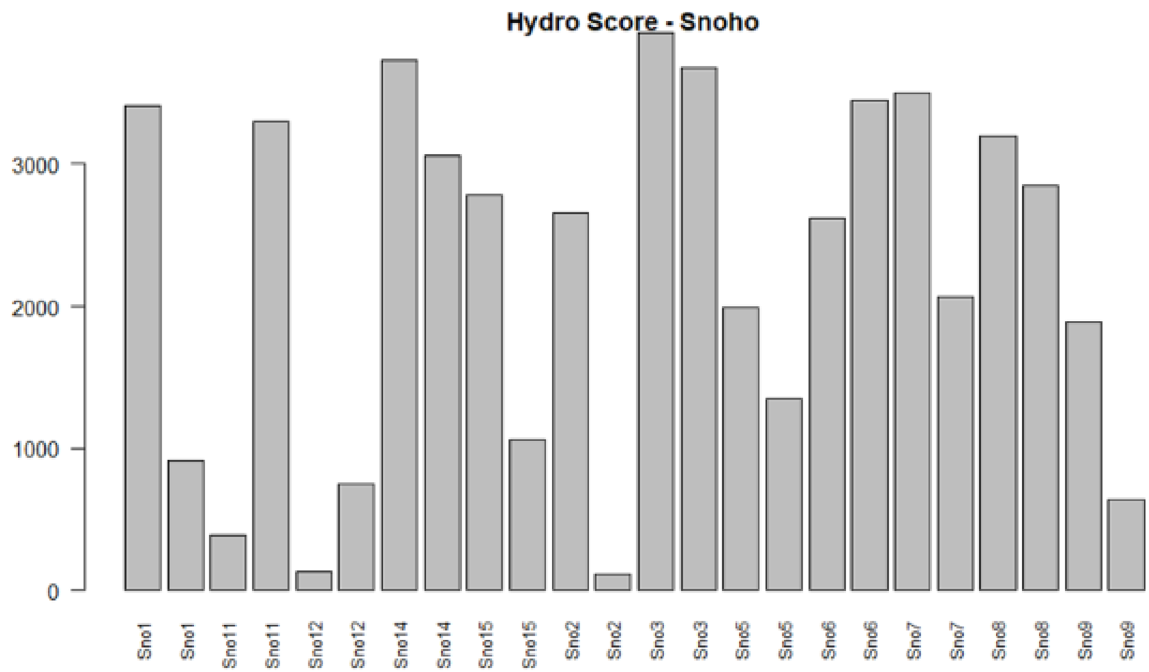
4. Effectiveness findings of rain gardens and bioretention facilities assessed

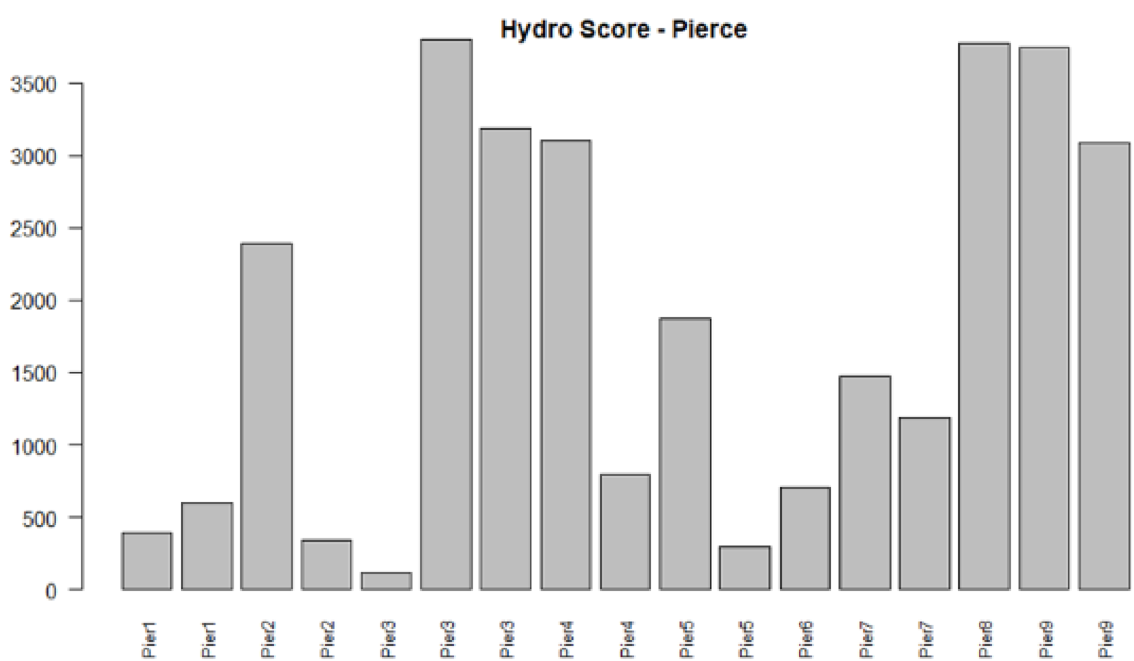
In an effort to extrapolate our findings to make broad generalizations about the overall effectiveness of rain gardens and bioretention facilities in terms of hydrological functions and vegetative functions, we created a hydrologic "score" and a vegetation "score" for each site assessment (Figure 6 a-d and Figure 7 a-d respectively). Using data collected from protocol version 2 implementation, hydrology and vegetation scores were calculated using an a priori assignment of weight to each metric and the range of possible values assigned to that metric. The weighting of values was based on our team's own experience and our review of the rain garden and bioretention monitoring research literature, as well as input from our Technical Advisory Committee.[2]

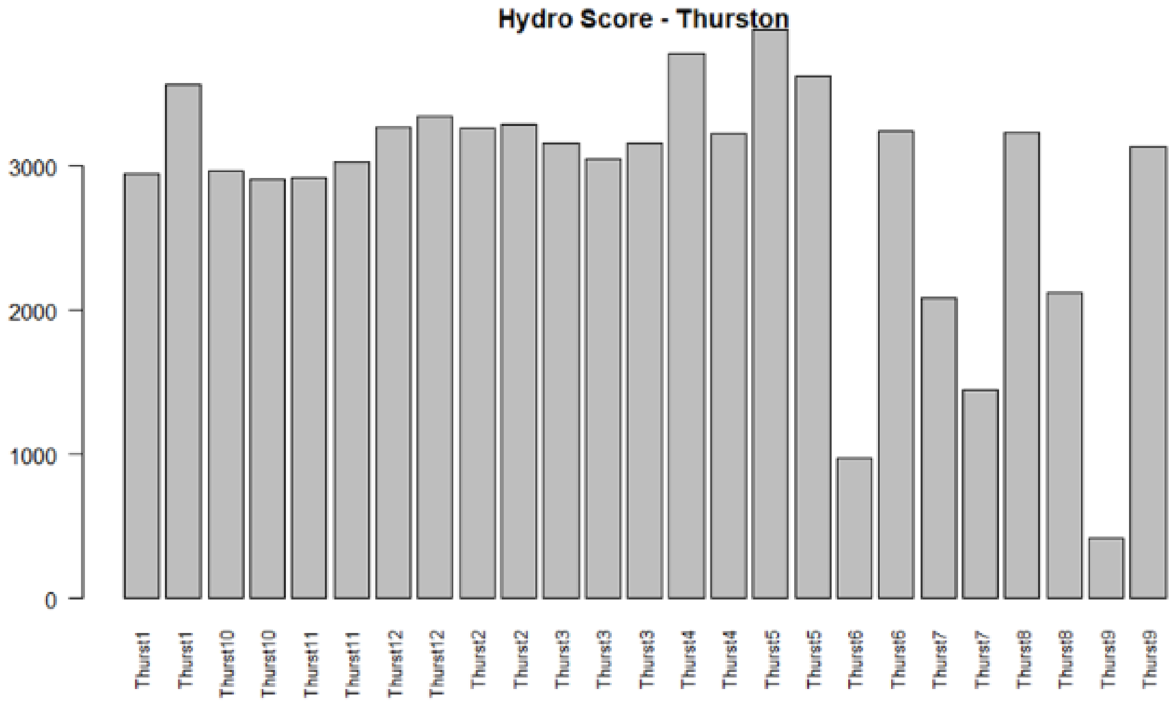
In this preliminary analysis we see a wide range of overall scores for different site assessments that received a full assessment by trained volunteers (n=47). 21 assessments scored a site below 2000 (an arbitrarily selected threshold for the sake of comparison) on this scale, and 6 sites received scores both above and below 2000. This analysis suggests that creating an overall score or grade for a single facility's hydrology based on observational data collected by volunteers on any given day (regardless of weather) may be informative, but more refinement to the formula is likely needed in order to increase the reproducibility of such scores across different assessors and different weather conditions.[2]

In the case of creating vegetative overall scores for each assessment (Figure 7 a-d), we found that scores varied even more widely than with the hydrologic scores. Some sites were found to have consistently low vegetation scores and a large number of sites had widely different scores when assessed by different teams on different days.[2]

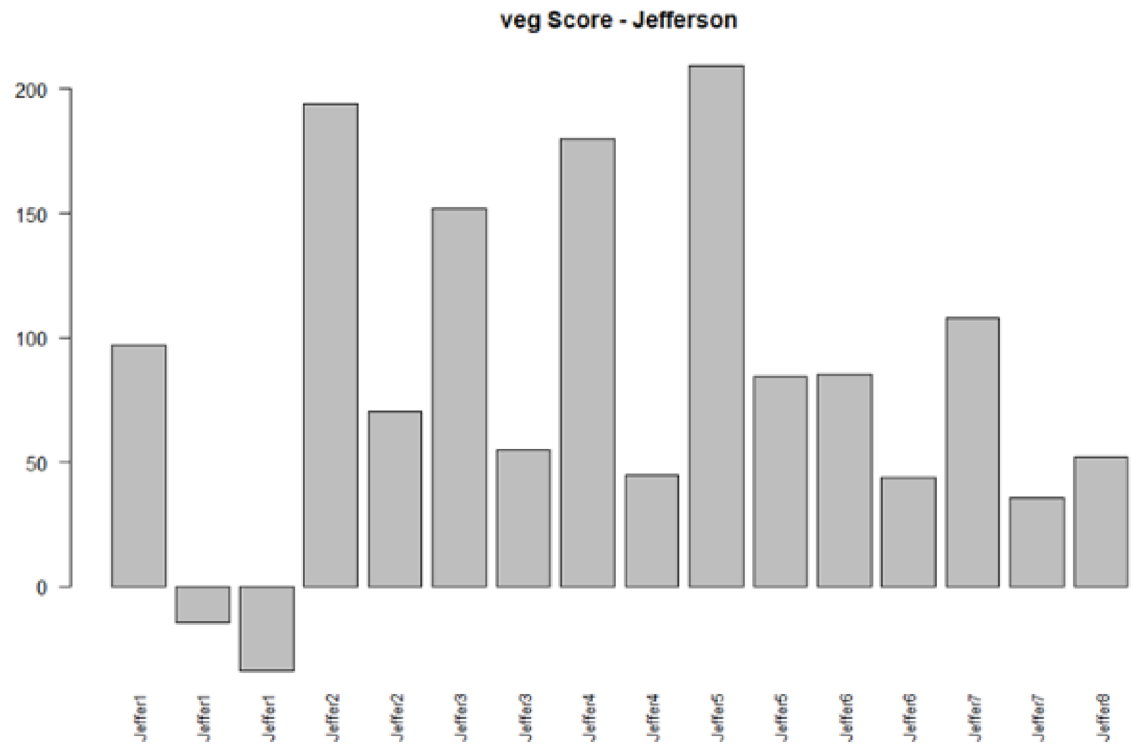


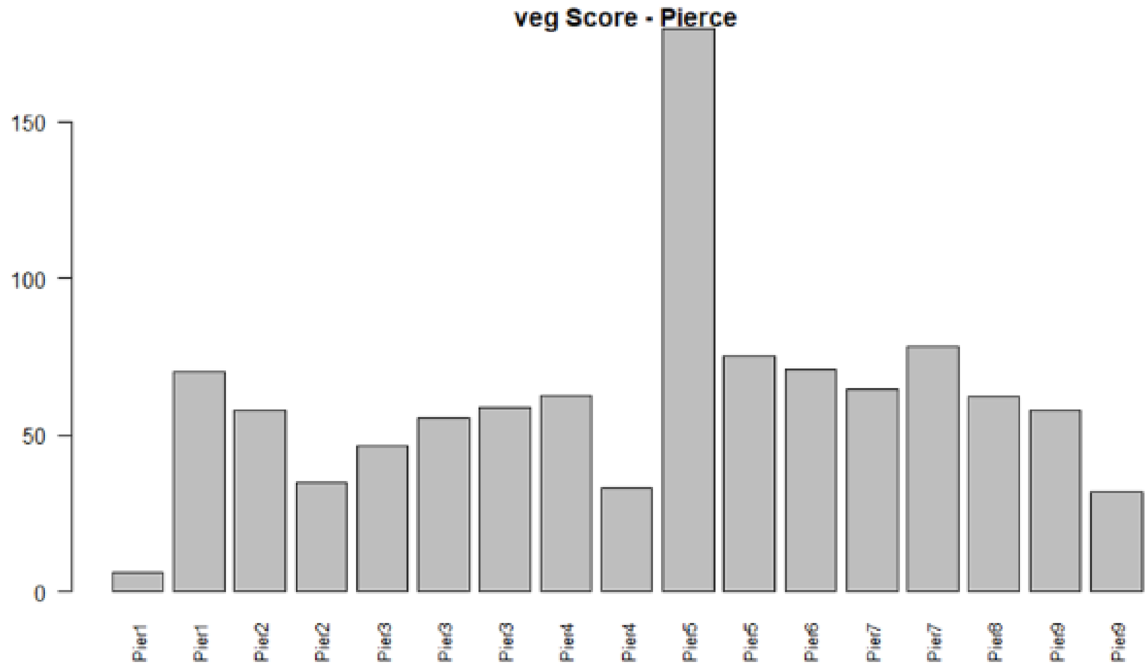

Hydro Score - Snoho
0
1000
2000
3000
Sno1
Sno1
Sno11
Sno11
Sno12
Sno12
Sno14
Sno14
Sno15
Sno15
Sno2
Sno2
Sno3
Sno3
Sno5
Sno5
Sno6
Sno6
Sno7
Sno7
Sno8
Sno8
Sno9
Sno9

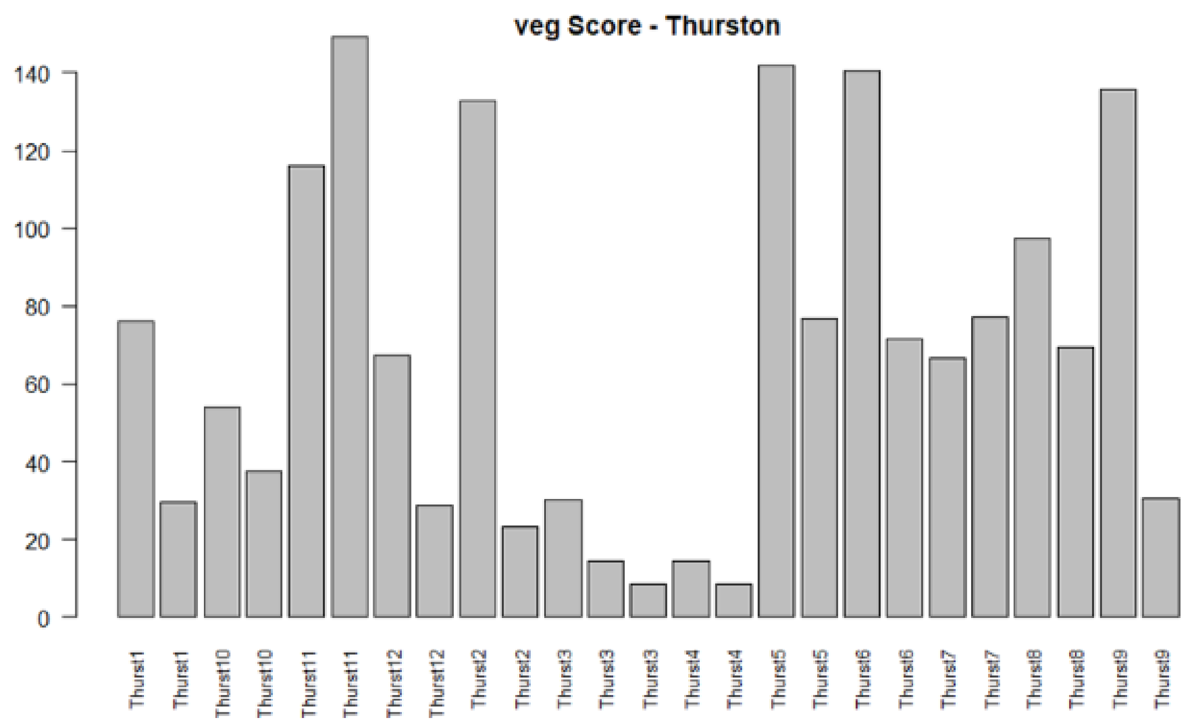

Hydro Score - Pierce
0
500
1000
1500
2000
2500
3000
3500
Pier1
Pier1
Pier2
Pier2
Pier3
Pier3
Pier3
Pier4
Pier4
Pier5
Pier5
Pier6
Pier7
Pier7
Pier8
Pier9
Pier9

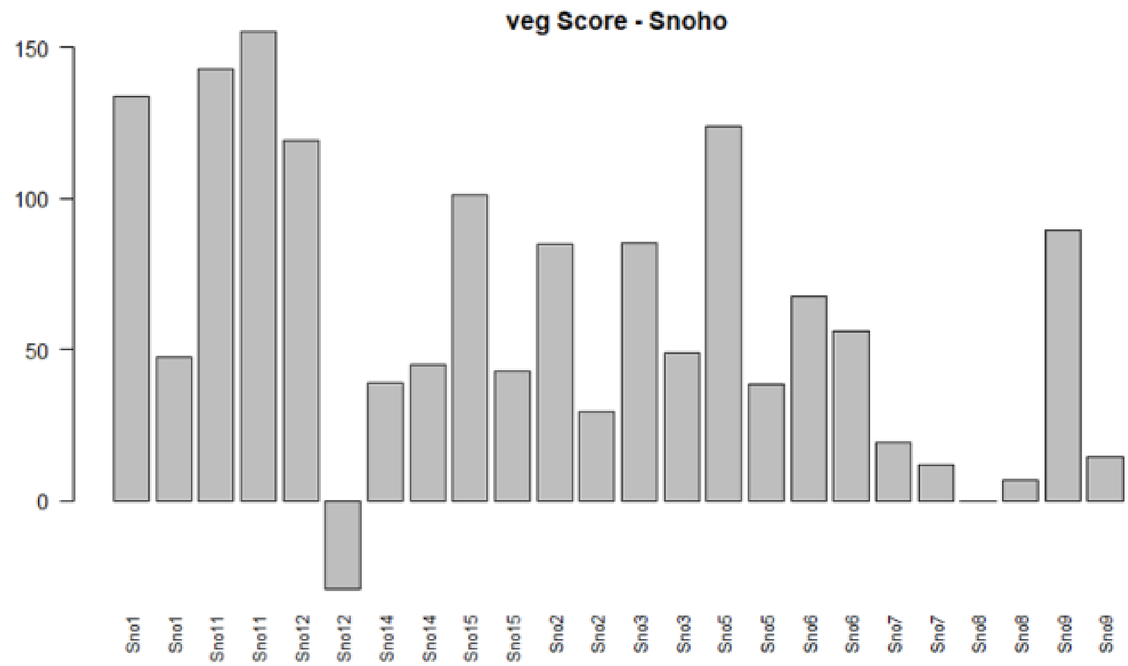


Figures 6a, 6b, 6c, and 6d: Transformed hydro scores by County. This score is based on a preliminary rubric that assigned values ranging from -100 to +100 for each metric in the assessment and then summed. Negative numbers were ascribed to metrics that were considered detrimental to rain garden functioning, like blocked inflows. Similarly, positive numbers were ascribed to metrics that were deemed beneficial to rain garden functioning. A full listing of how metrics were transformed can be made available. With the same rubric applied to all assessments, it is evident that Thurston and Jefferson sites were generally assessed to be in better condition than in Pierce and Snohomish where there is a lot more variability in scores, especially in the hydrologic functions.[2]


veg Score - Jefferson
200
150
100
50
0
Jeffer1
Jeffer1
Jeffer1
Jeffer2
Jeffer2
Jeffer3
Jeffer3
Jeffer4
Jeffer4
Jeffer5
Jeffer5
Jeffer6
Jeffer6
Jeffer7
Jeffer7
Jeffer8


veg Score - Pierce
150
100
50
0
Pier1
Pier1
Pier2
Pier2
Pier3
Pier3
Pier3
Pier4
Pier4
Pier5
Pier5
Pier6
Pier7
Pier7
Pier8
Pier9
Pier9


veg Score - Thurston
140
120
100
80
60
40
20
0
Thurst1
Thurst1
Thurst10
Thurst10
Thurst11
Thurst11
Thurst12
Thurst12
Thurst2
Thurst2
Thurst3
Thurst3
Thurst3
Thurst4
Thurst4
Thurst5
Thurst5
Thurst6
Thurst6
Thurst7
Thurst7
Thurst8
Thurst8
Thurst9
Thurst9



Figures 7a, 7b, 7c, and 7d: Transformed vegetative scores by County –Similar to Figure 4, these scores are based on a preliminary rubric that assigned values ranging from -100 to +100 for each metric in the assessment, and then summed. Negative numbers were ascribed to metrics that were considered detrimental to rain garden functioning, like the presence of weeds. Similarly, positive numbers were ascribed to metrics that were deemed beneficial to rain garden functioning. A full listing of how metrics were transformed can be made available. Negative scores are apparent at two sites in Jefferson and one in Snohomish – most likely due to a dominance of weeds or invasives.[2]

The TAC and other collaborators, suggested that we create an analysis process for assessment data that triggers “flags” when a single metric is scored above or below a certain critical threshold. A combination of metrics that fail to reach a threshold value could also be developed to raise a flag indicating that operations and maintenance attention is needed for that site in order to restore its hydrologic or vegetative functions. For some jurisdictions and agencies, it may be the case that only hydrologic flags are prioritized, while others may choose to attend to flags raised due to hydrological or vegetative scores. [2]

Based on this input from the TAC, the project team identified 57 parameters in the version 2 assessment protocol that were important to facility function as they relate to the three main assessment categories (hydrology, vegetation, and social aspects). The project team assigned levels of concern (flags) from moderate (yellow) to high (red) (See Appendix 6 for recommended levels of concern). Appendix 7 provides an analysis of version 2 data using the levels of concern recommended in Appendix 6.

The analysis of the version 2 data identified at least one red flag at all 47 sites, with a mean of 12.8 and a median value of 12 red flags per site. Interestingly, the analysis showed yellow flags only in 27 of the assessments, with a mean of 1.4 and a median value of 1 yellow flag per site. Tables 3 and 4 show the parameters which were of high concern (red flagged) at greater than 10% of the sites.

Table 3: Hydrology parameters listed in order of frequency that concern levels were reached in 47 sites assessed.

Hydrology Parameters	# of Red Flags	% Sites Red flagged
Mulch depth zone 1A	45	96%
Mulch depth zone 1C	43	91%
Mulch depth zone 2	43	91%
Mulch depth zone 1B	41	87%
Mulch depth zone 3	41	87%
Cover bare ground zone 1	9	19%
1 Blocked inflow	7	15%
Cover bare ground zone 3	7	15%
Cover bare ground zone 2	6	13%
Overflow concerns	5	11%
2 Blocked inflows	5	11%
Blocked sheet inflow	5	11%

Table 4: Vegetation parameters listed in order of frequency that concern levels were reached in 47 assessments.

Vegetation Parameters	# of Red Flags	% Sites Red flagged
Evergreen plant vigor zone 1	35	74%
Groundcover plant vigor zone 1	34	72%
Groundcover plant vigor zone 2	27	57%
Evergreen plant vigor zone 2	22	47%
Herbaceous plant vigor zone 3	20	43%
Herbaceous plant vigor zone 1	19	40%
Groundcover plant vigor zone 3	19	40%
Deciduous plant vigor zone 1	17	36%
Herbaceous plant vigor zone 2	17	36%
Evergreen plant vigor zone 3	15	32%
Deciduous plant vigor zone 2	14	30%
Problem plants zone 2	7	15%
Problem plants zone 3	7	15%
Problem plants zone 1	5	11%

It would be determined by the ultimate user of the assessment whether or not to further investigate and potentially provide remedial action. Considerations would include the specific parameter identified, seasonality, time and staffing constraints. This information could also be used to improve future facility maintenance or design. An example of a potential design change could be informed by the lack of vigor of evergreen plants in zone 1 that was demonstrated in 35 of the 47 sites assessed. Another example is the lack of adequate mulching (defined as none or trace amounts) found in all zones in a predominance of the sites. The user of that information may be mostly concerned with hydrologic functionality and may not make it a priority to address that issue, however they may more readily address issues such as blockages that directly impact the immediate effectiveness of the facility.

For efficiency, maintenance actions like clearing an inflow or overflow could be integrated into the assessment process itself, integrating the assessment and maintenance actions into a single site visit. This integration could be done by adding maintenance tasks to the assessment process or adding assessment to the maintenance process (within the acceptable labor rules of all parties).

Each parameter in version 2 of the protocol was evaluated by the project team for suitability and potential changes for inclusion in the final protocol. Appendix 8 provides spreadsheet detailing the project team notes for each parameter from version 2.

5. Summary of findings and recommendations:

- These results demonstrate that the assessment protocol can provide an overall indication of a rain garden or bioretention facility's current state and inform the appropriate maintenance action needed to restore effectiveness of the facility (e.g. replenish mulch or schedule regular blockage checks of the inflow).
- The protocol provides replicable results, as shown by the relatively strong consistency in assessments of the same site between different implementers.
- Extensive training of assessment implementers is not necessary, at least for the rapid assessment protocol, as shown by the relative consistency in assessments between trained and untrained implementers. Untrained volunteers did not implement the full protocol.
- The rapid assessment should provide the level of detail necessary to indicate if further actions are needed at a site, as well as to provide direction for future maintenance and some design considerations.
- The full assessment protocol, especially the vegetation metrics in version 1 would inform future design considerations and long term management strategies. Thus, that aspect of protocol version 1 may be useful to retain and borrow from as an assessment tool for specific research and design goals.
- Photo points utilized in version 1 would provide long term records and may provide some clarifying information on specific concerns for asset management staff without

requiring an additional site visit, however is not necessary for basic assessment purposes.
- Vegetation assessment information can be variable between assessment implementers and seasons. Therefore, while informative, vegetation data should be interpreted with an awareness of its inherent variability under this protocol.

The recommendations provided in appendix 8 will be utilized to finalize the protocol, which is deliverable 2.6 of this project. Information from this report will be utilized to outline the next steps in the "Proposal for a scaled-up monitoring program" which is deliverable 3.2b of this project.

Citations:

1. Bertolotto, Chrys and Aaron Clark. 2017. **Rain Garden and Bioretention Literature Review: An Assessment of Functional Parameters, BMPs and Landowner Perspectives**. Deliverable 1.1, Rain Garden and Bioretention Assessment Protocol Project. WA Department of Ecology, Stormwater Action Monitoring Program. https://ecology.wa.gov/Regulations-Permits/Reporting-requirements/Stormwater-monitoring/Stormwater-Action-Monitoring/SAM-effectiveness-studies/Raingarden-and-bioretention

2. Clark, Aaron. 2018. **Technical Memo: Results from Version 2 Assessment Protocol Implementation**. Deliverable 2.3, Rain Garden and Bioretention Assessment Protocol Project. WA Department of Ecology, Stormwater Action Monitoring Program. https://ecology.wa.gov/Regulations-Permits/Reporting-requirements/Stormwater-monitoring/Stormwater-Action-Monitoring/SAM-effectiveness-studies/Raingarden-and-bioretention

3. **Technical Advisory Committee (TAC)** for this project consisted of:
   Apryl Hynes (City of Everett); Leska Fore (Puget Sound Partnership); Ben Alexander (Sound Native Plants); Melissa Buckingham (Pierce Conservation District); Cari Simpson (Urban System Design); Mieke Hoppin (City of Tacoma); Curtis Hinman (Herrera Inc.); Doug Hutchinson (City of Seattle); with assistance from Brandi Lubliner (WA Ecology)