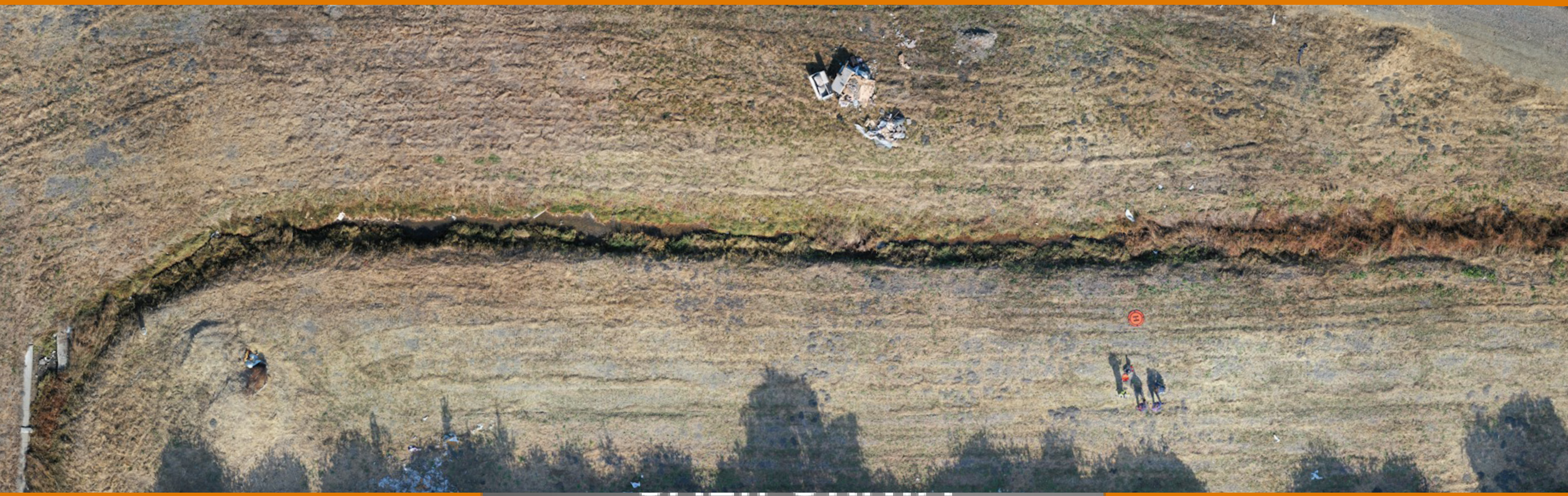
SCCWRP #1190

Field Testing Report

for the Statewide Trash Monitoring Methods Project

12.20.2020



FUNDED BY

California Ocean Protection Council

PREPARED BY

Southern California Coastal Water Research Project
3535 Harbor Blvd., Suite 110
Costa Mesa, CA 92626

AUTHORS

Shelly Moore, Tony Hale,
Stephen B. Weisberg, Lorenzo Flores,
Pete Kauhanen

San Francisco Estuary Institute
4911 Central Ave.
Richmond, CA 94804

FIELD TESTING REPORT FOR THE STATEWIDE TRASH MONITORING METHODS PROJECT

Contact	Role	Phone/Email	Report Approved
Tony Hale	Co-PI	tonyh@sfei.org	TH
Shelly Moore	Co-PI	shellym@sfei.org	SM
Pete Kauhanen	Researcher	petek@sfei.org	PK
Lorenzo Flores	Researcher	lorenzof@sfei.org	LF

Prepared by

Tony Hale, PhD

Shelly Moore

Stephen B. Weisberg

Lorenzo Flores, MS

Pete Kauhanen, MA

Table of Contents

List of Figures

List of Tables

Executive Summary

Trash has received renewed focus in recent years as policy makers, public agencies, environmental organizations, and community groups have taken many steps towards trash quantification and management across California. The range of management actions is matched by the diversity of monitoring approaches, designed to determine key attributes associated with trash pollution on California's lands and in its waterways.

This report describes the field testing associated with a project designed to validate the accuracy, precision, and practicality of several trash monitoring methods, practiced across the state. Additionally, the project measured the efficacy of a novel monitoring method designed to detect trash via remote sensing and machine learning.

In this report, readers will find details about each respective method -- the specific approach to landscape characterization, the qualitative or quantitative measures undertaken, the team-based quality assurance for data collection -- as well as the approach that the testing team adopted to ensure efficient, accurate, and useful validation of the methods.

Accordingly, because the validation efforts integrated multiple methods, using multiple teams at a selection of common sites, the field testing report yields useful statistical information not only about each method individually, but about the comparability of the results. The report illustrates the correlation factor associated with different forms of trash metrics, associated with different methods practiced on the same assessment sites. The results illustrated a generally high degree of correlation among different methods, which promises opportunities to compare results meaningfully across methods.

Furthermore, this field testing report provides quantitative measures to illustrate the repeatability of each method, the differences and insights yielded by assessment site sizing criteria varying among methods, the transferability / teach-ability of each method among trash monitoring practitioners, and how the degrees of accuracy might aid programs in performing mass balance analysis of known sources to trash detected in a given site.

Regarding innovation, the project team leveraged multiple on-the-ground methods and special testing scenarios to compare conventional and novel (aerial) assessments to measure the relative accuracy and precision of this emergent technology that might address some of the resource constraints that currently limit the broader or more frequent deployment of conventional trash assessment methods.

The analyses captured in this field testing report offer specific quantitative measures of the accuracy (bias), precision (repeatability), practicality and cost associated with each method. This information is subsequently used to inform a companion summary analysis found in the Trash Monitoring Playbook, which is designed to evaluate the applicability of the monitoring methods to address classes of monitoring questions.

Introduction and Overview

Trash on land has recently become a focus of policy throughout the state of California. These policies include three main areas: 1) bans; 2) total maximum daily loads (TMDLs); and 3) the Statewide Trash Amendments. While these policies all involve reducing trash on land, they all work at distinct levels. Bans typically target specific items such as the statewide ban on plastic bags. Local product bans and use restrictions throughout the state include specific items such as polystyrene and cigarettes. As mandates from regional water boards, TMDLs have been passed across the state on many contaminants and specifically for trash in at least 15 water bodies. The most well-known TMDL for the Los Angeles River was one of the nation's first trash TMDLs and was established in 2001. The goal of 100% trash load reduction for this TMDL was set to be accomplished by September 2016. Many jurisdictions have reached this using full trash capture systems or alternative institutional controls such as street sweeping, education, etc. The Statewide Trash Amendments take the TMDLs to a broader level, as jurisdictions throughout the state now must either install full trash capture devices (Track 1) or partial capture devices and institutional controls (Track 2). For those opting for Track 2, monitoring is required to ensure they are attaining results comparable to Track 1 areas.

Parallel to these regulatory and legislated actions are community-based organizations motivated to monitor and often remediate areas impacted by trash. These organizations often serve missions of environmental stewardship, community engagement, municipal beautification, economic development, recreational opportunity, or any combination of the above. Trash monitoring practitioners, ranging from weekend volunteers to vocational professionals, commit significant time to measuring the load, status, sources, and impacts of trash in various habitats across California.

In April of 2017, stakeholders involved in ensuring the health of our environments through the control of escaped trash, met in Oakland, California to discuss recent trash policies, assess monitoring their effectiveness, and share their concerns regarding ongoing and future monitoring assessments. What came out of the meeting was a recommendation for a study to determine methods to assess trash in rivers and streams that could be used to provide information on both temporal and spatial scales. This document represents the testing of methods to assess trash in rivers and streams, to determine the management and scientific questions these methods can assess for and the resources needed, as well as the strengths and weaknesses of these methods for repeatability and accuracy.

Multiple agencies across California have developed standardized methodologies for monitoring trash in the environment. As municipalities and water-quality regulatory agencies have implemented programs and policies to more effectively manage and control trash loading to storm drain conveyances, there has been increased interest in using these methods to quantify the effectiveness of management actions. To create a foundation for developing a consistent, standardized approach to trash monitoring statewide, this project assessed the accuracy, repeatability, and efficiency of these standardized trash monitoring methodologies already in use, as well as investigated a new, innovative method. Methods developed by the Bay Area Stormwater Management Agencies Association (BASMAA) for use in the San Francisco Bay Area were compared to methods developed by the Southern California Stormwater Monitoring Coalition (SMC) for use in coastal southern California. These methods were selected for their broad influence in their respective regions, on the one hand, and their divergent approaches to trash monitoring, on the other. One of the chief goals of this project was to understand the similarities and differences among the already existing methods for detecting, quantifying, and characterizing trash in selected environments.

The findings of this project will be used to inform a statewide effort to develop rigorous, monitoring methods to support the State Water Board's Trash Amendments, which over the next couple of decades will require all California dischargers in high trash generation areas to either install full-capture devices at storm drain inlets or develop a plan to capture or reduce the amount of escaped trash at equivalent rates. This report specifically details the results from comparing three already established methods and one novel method. The end goal is to produce a Playbook of methodologies that stakeholders can consult to determine the best method to use based on their specific management questions, the amount of resources they have, and the given method's performance compared to others.

Project Context

Regulatory / Mandate Criteria

The Statewide Trash Amendments give jurisdictions the option of choosing Track 1 (full capture devices) or Track 2 (partial capture devices and institutional controls) for reducing trash. Those opting for Track 2 must develop and implement a set of monitoring objectives that demonstrate effectiveness of the selected combination of controls and compliance with full capture system equivalency. This monitoring must indicate compliance comparable to Track 1. There are many other methods that can be used to show full capture equivalency (e.g. On Land Visual Assessment). The methods in this document are specific for estimating trash in receiving waters only and are therefore only minimally effective in determining pathways-specific contributions to the receiving waters.

Problem(s) to Address

While sampling is taking place in receiving waters (BASMAA 2017; Moore et al. 2016), there are no accepted standardized methods to address the amounts of trash. In addition, methods that can provide both spatial and temporal comparisons in a variety of habitats are needed. There is a demonstrated need for method options based on a variety of factors, such as the monitoring questions posed, method bias, repeatability, practicality, and available resources.

Purpose of Testing

Several different monitoring methods have emerged within the context of various programs throughout the state. The methods either address the same management questions using different approaches, or different management questions altogether.

Three already established methods were tested for this project. These include methods developed by both the Bay Area Stormwater Management Agencies Association (BASMAA) in the San Francisco Bay Area and the Stormwater Monitoring Coalition (SMC) in Southern California. Taken together, these methods offer useful contrasts in approaches, including different regional influences, different regulatory approaches, and different habitats.

This project conducted field testing of the methods to assess comparability in three main respects:

- To what extent are the methods cross-comparable in the outputs they yield?
- To what extent do the methods represent accurate measures of trash in the environment?
- To what extent is each method internally consistent to yield similar results by different practitioners?

This project also evaluated the practicality and applicability of novel monitoring approaches, including the use of unoccupied aerial systems (UAS) and machine learning algorithms to aid with the detection, quantification, and characterization of trash in selected environments. The project assessed the utility of these methods in the context of existing monitoring approaches.

Validation of Current Methods

Chief among the objectives of the field testing were to validate the methods as performed by the BASMAA and SMC practitioners. The methods were validated in terms of their precision (repeatability) and accuracy (bias), as well as evaluating the amount of resources needed (Table 1). Precision measures the amount of variation in results among practitioners of the same method or using the same instrument, whereas accuracy measures the degree to which a practitioner's results vary from whatever is determined to be the "true" value.

Both factors are critically important to measure when establishing meaningful, practical standards. A summary of the comparison matrix used for this study indicates all of the parameters looked at, as well as what types of monitoring questions might be answered by each method (Table 1).

Table 1. Comparison matrix for tested methods.

METHOD	MONITORING QUESTIONS	ACCURACY	REPEATABILITY	RESOURCES
A				👤👤👤👤
B				👤👤👤
C				👤👤
D				👤

Monitoring Questions

The type of monitoring questions answered by each method will determine the method to be used by a given stakeholder. Some may be interested in only assessing the general condition of the site based on a subjective view of the amounts of trash present, while others may want more specific information relative to the amount of space (volume) that is occupied by the trash or the number of specific items, such as cigarettes or plastic bags found in the environment. Understanding the management question is essential to determining the monitoring questions.

Accuracy / Bias

Accuracy of the methods were measured by comparing survey results to the trash that was extracted from the assessment area. Extraction is the operative measure of truth. The team used both tally and volumetric measures to assess the amount of trash.

Precision / Repeatability

Precision of the available methods was determined by conducting multiple assessments of the same site, under the same conditions, using the equivalent instruments, with different teams. The amount of variability among the results is inversely proportional to the degree of precision.

Resources

Resources are often a limiting factor in performing assessments. One of the most mentioned issues when talking about trash assessments is the amount of time it takes to do an assessment. This project measured the times needed to do each assessment to provide estimates of how long a particular assessment would take.

Assessment of UAS Methods

The project relied on a slightly different approach to assess the repeatability and accuracy of novel methods. Because the necessary data collection resources (i.e., unoccupied aerial vehicles) are specialized and costly, having multiple teams repeat the data collection in the field would be impractical. Such an approach would also fail to test the actual assessment / analysis primarily under consideration. Instead, we concentrated on testing the repeatability of the data collection under varying conditions at different sites, paired with the repeatability and accuracy of the data analysis it facilitates.

Precision / Repeatability

In using UAS for data collection, we repeated the surveys in different assessment areas, under different conditions, using the same instruments. This project measured the repeatability of the sensors used to detect trash in the environment.

High degrees of variability of the information collected would indicate that the instruments are intrinsically imprecise or there are external environmental conditions that lead to occlusion of the optical information. Analyzing the “outliers” in the results helped to determine the source of the variability.

Testing the repeatability of the analysis is important, since, in the case of the remote-sensing methods, the assessment occurs not on-site but rather later at a desk. Therefore, in testing the precision of the observation method, multiple parties were asked to identify trash in the resulting imagery.

Accuracy / Bias

By pairing UAS-based assessments with conventional, extraction-based assessment events, the team was able to measure the accuracy of its imagery-based results. Without ground-truthing, the remote sensing would be subject to limitations of image resolution and other factors affecting visibility. However, the well-timed on-site assessments provide the “truth” values for comparison.

Practicality

With the novel methods, we assessed whether the proposed methods were practical. In this context, practicality is measured by:

- ease of use
- speed of assessment execution
- cost
- overall effectiveness

Comparability of Results

Because different methods have been established in Southern and Northern California, one of the important outputs of this project was to measure how the values emerging from the different methods could be effectively cross-walked from one to the other. In other words, how did counts translate to volumes? How did condition assessments translate to counts or volumes? How did remote sensing methods bridge several methods?

The team used volumes instead of weights as a common form of comparison because 1) weights are more variable in riverine environments where inundation and sediment fouling are very common, 2) weights are less meaningful in terms of understanding the aesthetic impacts of trash in the environment, and 3) weights fail to capture the enduring impacts of lightweight plastics on wildlife and receiving water ecosystems.

If credible cross-walking formulae can be established, the results from each method, practiced across the state, would be brought together in some fashion, with known limitations, constraints, accuracy, and precision appropriately identified.

Testing Phases

To optimize the level of effort and promote the best deployment of its instruments, the team undertook the field testing in several phases, as adumbrated in Figure 1.

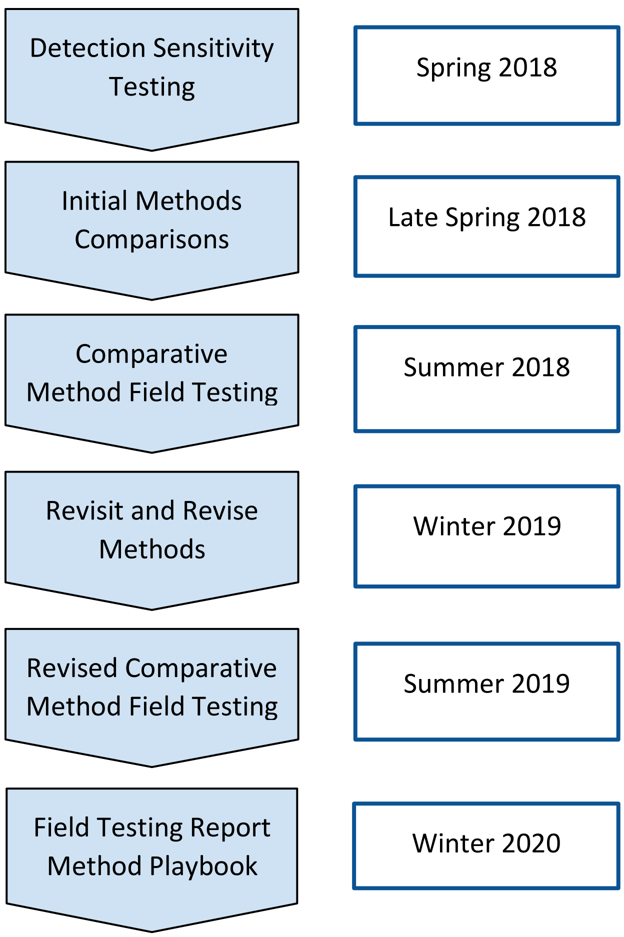


Figure 1. General schedule for field testing through 2020.

The first phase took place in spring 2018 and focused on determining the detectability of trash relative to various site conditions and technical instruments. During this phase, the team received initial training on the in-stream assessment methods used for this project and began performing some UAS-based surveys as part of a series of instrumentation tests.

Because different methods have been established in Southern and Northern California, one of the important outputs of this project was to measure how the values emerging from the different methods could be effectively cross-walked from one to the other.

Next, in late spring, the team made some comparisons of the methods it has been practicing, and made initial judgments about their cross-comparability, ambiguity factors, and accuracy.

The primary testing phase began in summer 2018; the team coordinated their surveys with the BASMAA and SMC teams.

Following the survey season, the analysis of the results was begun with full intensity. The cross-comparability of the results were measured, and the development of the machine learning algorithm was intensified. The team also undertook any revisions to the testing approach.

In summer 2019, the next round of surveys began. The team focused on assessing 8 additional sites each in Northern California and Southern California with larger amounts of trash and on incorporating the Escaped Trash Assessment Protocol (ETAP) trash assessment method developed by the United States Environmental Protection Agency (USEPA).

Coordination with Other Projects

To assess methods for trash assessments in California, the Project Team reached out to others already performing trash assessments. Both the Bay Area Stormwater Management Agency Association and the Southern California Stormwater Monitoring Coalition were already performing two different types of assessments in their own regions. Additionally, the USEPA has developed a method similar to a synthesis of the SMC and BASMAA methods. This method, known as the Escaped Trash Assessment Protocol (ETAP), has not been finalized and changes are still being made to refine it. Below is a brief description of each agency and their trash methodologies.

Bay Area Stormwater Management Agency Association

The Bay Area Stormwater Management Agencies Association is a non-profit organization made up of municipal stormwater programs in the San Francisco Bay Area. Their purpose is to coordinate and facilitate regional activities for the municipal stormwater programs, focusing on regional challenges and collaborative opportunities to meet stormwater program requirements. In 2017 they developed the Receiving Water Trash Monitoring Program Plan for the San Francisco Bay Region (BASMAA 2018) through collaboration with regional stakeholders and scientific peer reviewers. They have also produced the Standard Operating Procedures and Data Collection Forms for Qualitative Trash Assessments and Quantitative Trash Monitoring in Receiving Waters (BASMAA 2020), delineating trash protocols for Bay Area stormwater agencies to use to assess trash. Most recently they are finalizing the San Francisco Bay

Area Receiving Water Trash Monitoring: Pilot-Testing of Qualitative and Quantitative Monitoring and Assessment Protocols, a study done to assess the monitoring methods they have developed and recommend in BASMAA 2020. Methods used by BASMAA to assess trash conditions are included in this report. Those methods include both Qualitative and Quantitative assessments through assessing trash levels visually and volumetrically. Additionally, estimates of source are provided as part of their volumetric assessment of trash. Their methods have been used in the assessment of methods included in this report.

Southern California Stormwater Monitoring Coalition

The Southern California Stormwater Monitoring Coalition (SMC) was formed in 2001 by a cooperative agreement between stormwater agencies in Southern California. Their goal through this collaborative effort and sharing of resources is to develop the technical information necessary to better understand stormwater mechanisms and impacts and the tools that will effectively and efficiently improve stormwater decision-making. The SMC develops and funds cooperative projects to improve the knowledge of stormwater quality management and reports on the progress of those projects on an annual basis. One such project is the Southern California Stream Survey, which is performed on an annual basis to assess the condition of rivers and streams. As part of this survey, they have included trash assessments in 2011-2013 and in 2018-2019. They performed both qualitative and quantitative surveys as well, using the same method BASMAA used to visually assess a site. Their quantitative method is a tally method that identifies and counts the number of trash items in a given area. Their tally method has been used to assess methods included in this report.

USEPA Escaped Trash Assessment Protocol

The Escaped Trash Assessment Protocol (ETAP) was created by the Trash Free Waters Program (TFW) of the United States Environmental Protection Agency (U.S. EPA). TFW works to reduce the amount of trash entering our waterways. Their goal is to develop a universal trash assessment method, applicable to all environments, that assesses trash that escapes the collection systems and ends up in waterways via transport mechanisms such as wind, storm drains, and improper disposal. Ultimately, they are looking to identify sources and implement effective prevention strategies in order to reduce the input of trash into our inland waters and the ocean. Their method not only counts items within an assessment area, but also assesses each item for hazard potential. This project assessed a subset of sites during our second field testing season to test this method and compare it to the primary methods assessed in this study.

Technical Advisory Committee (TAC)

A Technical Advisory Committee was brought together, consisting of trash method and policy experts. The goal of the TAC was to provide technical advice on the approach to the project, provide technical review of the products, and by ambassadors to the monitoring community. The TAC included a variety of subject matter experts, ranging from those in the field to those creating the policies to control trash in the environment.

In-Stream Monitoring Methods and Project Coordination

- Chris Sommers, EOA, Inc.
- Donna Bodine, GeoSyntec (through the first year)
- Ted Von Bitner, AMEC Foster Wheeler
- Karin Wisenbaker, Aquatic Bioassay Consultants

- Kate Wing, Kate Wing Consulting

Program-based Methods Development and Testing

- George Leonard, Ocean Conservancy
- Sherry Lippiatt, NOAA
- Kaitlyn Kalua, California Coastkeeper Alliance

Policy Makers, Funders

- Holly Wyer, Ocean Protection Council
- Greg Gearheart, State Water Resources Control Board

Field Methods Tested

Four field methods were tested for this project that answer a variety of management questions and cover diverse habitats within receiving water, rivers, and streams. The four methods are:

1. Qualitative
   a. **Visual** – this method involves walking the survey area, observing trash levels and subjectively assigning a score (1-12, with 1 being the lowest and 12 the highest) based on the perceived amount of trash found.
2. Quantitative
   a. **Trash Tally** – trash within the assessment area is identified and categorized based on type of material and use. A total count of trash, as well as counts of each type, are produced.
   b. **Volumetric** – trash within the assessment area is collected, placed in buckets and the volume is estimated and recorded by each trash category and overall.
3. Novel
   a. **Unoccupied Aerial System (UAS)** – Drone flights were conducted over an assessment area to produce imagery, which was analyzed both visually and through machine learning.

An additional method, not originally included in this project, was assessed during the second field testing season – see the Escaped Trash Assessment Protocol above.

Conceptual Approach

All four methods selected for testing were performed at receiving water sites in both Northern and Southern California to compare results. To leverage existing monitoring efforts, BASMAA and the Southern California Stormwater Monitoring Coalition (SMC) agreed to perform the project's testing methods while they were completing field work using their existing monitoring methods, and the full suite of methods were tested on a subset of the sites assessed by the SMC and BASMAA for this project. Both groups conducted trash surveys in both field seasons at sites starting in late spring/early summer and finished in early fall. While both groups performed visual surveys, each group performed only their own assessment methods. In other words, the Bay Area Stormwater monitoring agencies performed volumetric assessments, and the Southern California Stormwater monitoring agencies performed counts of trash items, as per their respective methodological approaches. Meanwhile, SCCWRP and SFEI project teams joined the BASMAA and SMC teams in the field at a subset of sites and performed additional surveys to ensure all assessment methods were performed at those sites, including the UAS method to produce imagery to assess for trash remotely. At sample sites where BASMAA and the SMC did not perform their methods, the sampling team utilized BASMAA and SMC trash protocols.

For all methods other than the UAS-based assessments, the teams consisted of a minimum of a pair of trained assessors. One performed the trash detection work while the other served as a scribe, taking notes of the assessment. When appropriate, the two assessors conferred to compare their judgment and arrive at a consensus.

For the UAS-based assessments, the team consisted of a pilot and a spotter, as per the FAA rules and recommendations.

Site Selection

Site selection is key to promote applicability of the findings to the broadest set of conditions. Therefore, at the recommendation of the project's Technical Advisory Committee (TAC), the team established the following criteria to target the factors that most commonly influence the outcomes of trash assessments.

Criteria

A minimum of 30 sites were targeted. The project team selected its sites according to a rubric that achieves diversity along two axes: trash level and vegetation level (Figure 2).

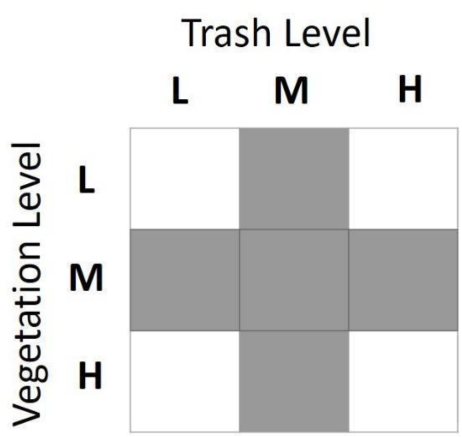


Figure 2. Target trash and vegetation levels (shaded areas) for test sites. (h=high, m=moderate, l=low)

The goal was to achieve a representative sample of sites that can validate currently practiced methods in high-, medium-, and low-trash areas in combination with areas of high, medium, and low levels of vegetation.

Regarding trash levels, the definitions of high, medium, and low generally correlate to the qualitative trash assessment categories established by BASMAA, with their associated quantitative estimates for gallons of trash per assessment site. (The quantitative relationship allows for an initial cross-walking of the BASMAA and SMC trash monitoring methods.) Low, by this logic, corresponds to the "low" condition category in our matrix. Moderate corresponds to the "moderate" category. And "high" to both the "high" and "very high" categories. The assessments account for items larger than 5mm.

The vegetation levels relate to the amount of vegetative cover, regardless of type. Knowing that different vegetation types interact with trash differently, our results also took into account the different types of vegetation, though we did not select our sites by these differences.

Because we are leveraging existing efforts, our pool of potential sites were largely driven by the opportunities presented by our BASMAA and SMC partners. Their combination of probabilistic and targeted sites formed the superset from which the project team selected its sites.

And because trash and vegetation levels can vary widely from site to site, depending on various factors, the team used satellite imagery from Planet.com (https://www.planet.com/) and aerial imagery from Google Earth and Google Maps to perform advance surveillance of sites immediately before a visit, assisting with determining vegetation levels and logistics. This offered additional screening value to ensure a representative diverse set of sites.

Please note that, while we could reasonably ascertain a predicted vegetation level, the team found it challenging to pre-determine trash levels, given that the levels were highly variable over time and space, based on a number of conditions including specific assessment boundaries, precipitation, wind, hydrology, and human intervention.

Locations

The conventional methods focus on creeks and streams. Accordingly, the sites were largely confined to these locations identified in the Bay Area, California Parks, and in Southern California.

However, for the novel methods, as the budget allowed, we conducted a limited number of surveys of wetlands located in Northern California in the Baylands around San Francisco Bay

For wetlands, particularly those that were inaccessible to foot traffic, the team had limited ability to ground-truth the remote-sensing imagery. There were no other assessments in these conducted, since neither the BASMAA nor SMC methods apply to vegetated wetlands. Despite this limitation, however, the team was still able to apply the same forms of analysis to the post-processed images, with the goal of prioritizing measures of precision over those of accuracy.

Vegetation Level \ Trash Level	L	M	H
L	12	4	3
M	2	4	2
H	9	8	2

Figure 3. Number of sites assessed in targeted areas with vegetation and trash levels, shaded for moderate conditions (H= High, M=Moderate, L=Low).

Sampling Success

A total of 50 sites over two sampling seasons were assessed for trash using the methods listed for this study (Figure 3). It was difficult to target sites directly based on the amount of trash and vegetation as in the first season the study team was pairing up with both BASMAA (in northern California) and the SMC (in southern California), to perform synchronous assessments at the same sites. In the second year, the Project Team targeted higher level trash sites and overall was able to assess sites ranging through all levels of trash and vegetation (Figure 4).

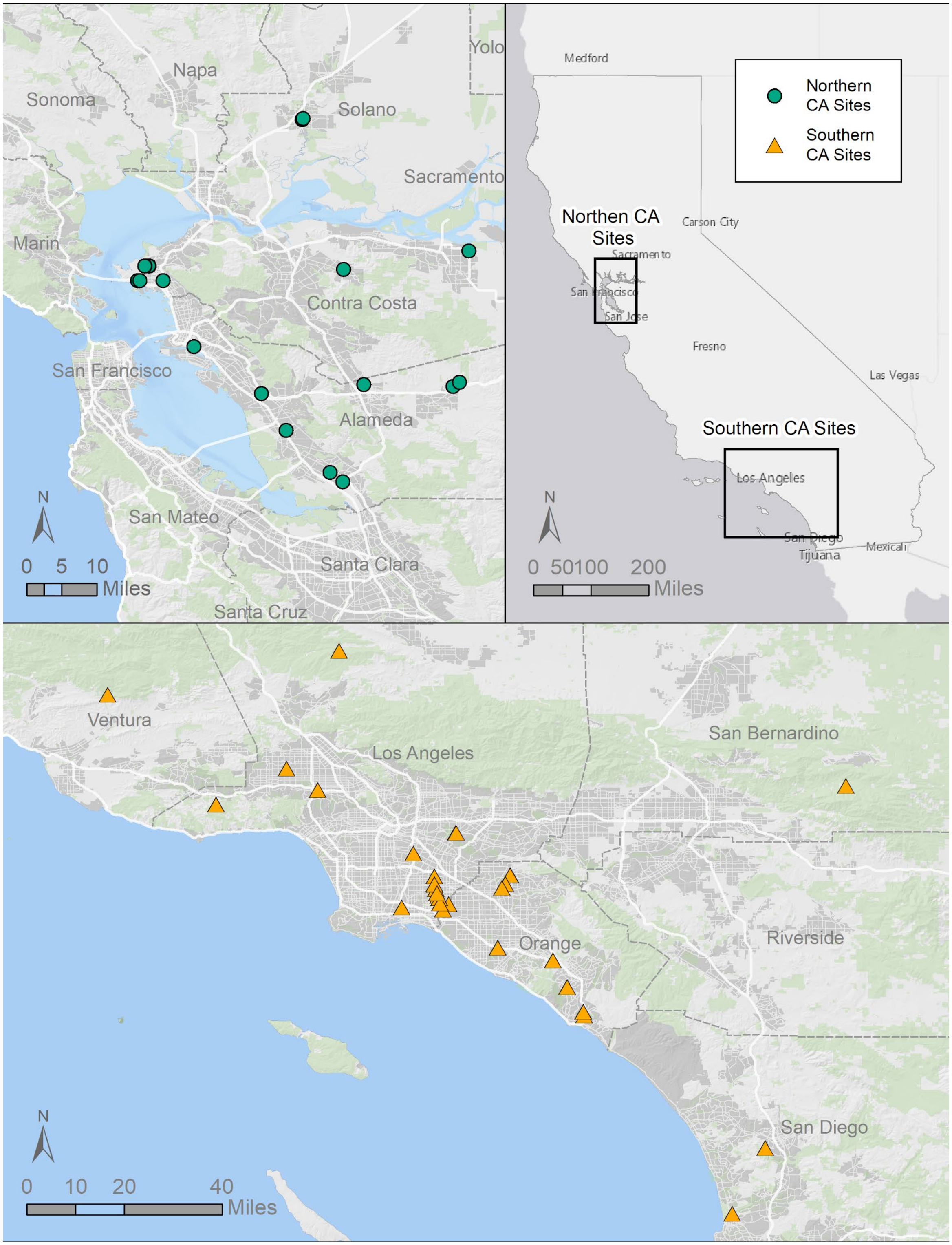


Figure 4. Location of sites in Northern and Southern California used to assess trash methods.

Site Information Assessment

Site information was collected at each site assessed for trash. This includes information on location, channel type, assessment area, proximity to stormwater drains and homeless encampments, and any other characteristics deemed important to the amounts of trash. This information was collected prior to beginning any of the trash assessments. It was important to collect this information as conditions can change over time and many of the site condition variables are thought to be important to understanding the amounts of trash residing in the assessment areas.

Sites included in this study were either probabilistic (as part of BASMAA's or SMC's survey sample draw) or targeted and included surveying 30- or 100-meter stream reaches. Before each trash assessment began, the monitoring reach was identified. For probabilistic surveys associated with the larger stream (bioassessment) surveys, Surface Water Ambient Monitoring Program (SWAMP) protocols were used (https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/), and the designated stream reach for trash assessment coincided with the first three transects A, B and C (Figure 5). SWAMP is a state-wide program that brings together comparable, high-quality data from carefully designed monitoring programs to provide information to resource managers, decision makers, and the public to assist with environmental management decisions.

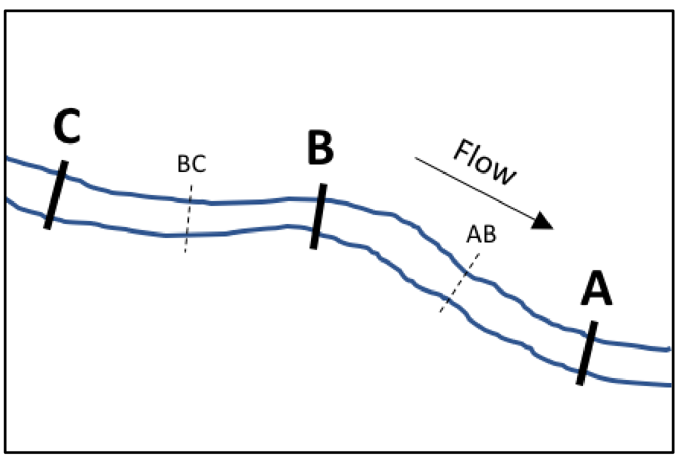


Figure 5. Diagram of transects A through C relative to stream flow for a Bioassessment sampling.[1]

Assessment Elements

The team completed the General Site Information on the trash survey field form (Figure 6). They recorded the Station ID, Start and Stop Time, Latitude and Longitude (in decimal degrees to at least 5 places), and the Datum (GIS projection used as a point of reference for the site locations). The survey

[1] Full SOPs can be found here: https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/combined_sop_2016.pdf

also documented members of the field crew conducting the survey and a brief River/Site Description and the Watershed Location of the site.

General Site Information

Station ID: ______________ Date: ______________

Start Time: ______________ End Time: ______________

StartLatitude: ______________ StartLongitude: ______________ Datum: ☐ NAD 83 ☐WGS 84

EndLatitude: ______________ EndLongitude: ______________ ☐ Other __________

Field Crew: ______________________________

____

River/Site Description: ______________________________ Watershed: ______________

Access:

Left Bank (circle one): Easy Moderate Hard Right Bank (circle one): Easy Moderate Hard

Channel Type (Check all that apply):

☐ Natural ☐ Earthen ☐ Concrete ☐ Rip Rap ☐ Other

Type of Site: ☐ Probabilistic ☐ Targeted Is stream flowing? Yes / No

Figure 6. General Site Information form.

Setting up the Assessment Area

The assessment area width extends to bankfull width of the stream (Figure 7). Bankfull width is determined by estimating the maximum water inundation in a one to two-year flood event (Ode et al., 2016; https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/combined_sop_2016.pdf). The team walked beyond the stream to look for evidence of one to two-year flood events. Evidence for bankfull locations include: topography, vegetation, sediment type, changes in bank slope and location of water stains on concrete or bedrock. Field crews viewed the video "A Guide for Field Identification of Bankfull Stage in the Western United States" (https://www.youtube.com/watch?v=UuS7H2NxJIM) and were part of training activities conducted by both BASMAA and the SMC. The team measured the wetted and bankfull width of the stream (Figure 8) using a measuring tape (a range finder may be used for larger streams) and recorded the measurement on the datasheet for each assessment site. Wetted and bankfull widths were measured at transects A (downstream extent), B (middle), and C (upstream extent). Teams also indicated whether trash was collected during the assessment.

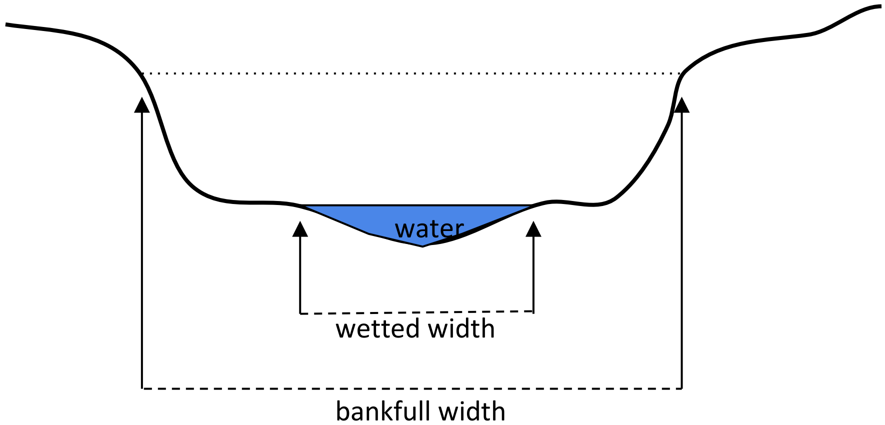


Figure 7. Stream cross sectional diagram of a typical stream channel showing the locations of wetted and bankfull width measurements.

Assessment Area						
Reach Length (m)	______________					
Wetted Width (m)	Transcect A	______________	Transect B	______________	Transect C	______________
Bankfull Width (m)	Transect A	______________	Transect B	______________	Transect C	______________
Trash picked-up during assessment? Yes / No						

Figure 8. Assessment Area form.

Stormwater Outfalls/Encampments

The team recorded the number and size of stormwater outfalls (greater than 18 inches) in the assessment area (Figure 9). Outfall categories are as follows: 18 - 24 inches; 25 - 36 inches; 37 - 48 inches; >48 inches. The team also recorded if there is trash at the outfalls and the amount of trash present. Trash amount categories were as follows: <10; <50; <100; >100 and represent the number of pieces present. The team determined if there was a homeless encampment in the assessment area or within 200 meters of the assessment area, and noted if the encampment was upstream, downstream, or within the assessment area.

Stormwater Outfalls/Encampments				
Number of stormwater outfalls in the assessment area >18" in diameter				
	18-24" ________	25-36" ________	37-48" ________	>48" ________
Trash at Outfalls? Yes / No				
Amount of Trash Present (number of pieces): (circle one)	<10	<50	<100	>100
Homeless encampment within 200 meters of assessment area? Yes / No				

Figure 9. Stormwater Outfalls/Encampments form.

Pictures

Trash conditions were photographed during each assessment. A minimum of 4 photographs were taken at each site at the beginning (upstream), middle (upstream and downstream), and end (downstream) of the assessment area (Figure 10). Additional photographs were taken to document site conditions.

Photo Documentation		
Segment	**Location**	**Photograph ID**
Bottom (A)	Upstream	
Middle (B)	Upstream	
	Downstream	
Top (C)	Downstream	
Other Photos	Misc. 1	
	Misc. 2	
	Misc. 3	

Figure 10. Photo documentation form.

Vegetated Condition Assessment

Two measurements were taken to estimate the amount of vegetation both in the water and on the banks at a given site (Figure 11). On the data collection form, the proportion (%) of the assessment area covered by vegetation or vegetative debris (i.e., large wood debris) was recorded for: 1) stream banks (combined area for both banks, including vegetated islands if present); and 2) the stream channel (wetted and/or dry). These measurements each added up to 100 percent and was used to determine how the vegetation relates to the amounts of trash found at a site.

Vegetated Condition Assessment				
▶ Below, Estimate the Proportion (%) of the total area of combined banks within the Assessment Area (including vegetated islands if present) that contains the following cover types:				
Ground Cover (e.g., grasses/weeds < 2ft in height)	**Understory** (e.g., bushes, poison oak, blackberries, small trees 2-10ft in height)	**Trees/Roots/Woot** (e.g., living trees/roots along toe of bank, other natural woody debris material)	**Bare Ground** (e.g., soil, concrete and other bank armoring material)	**Total**
___________%	___________%	___________%	___________%	**100%**
▶ Below, Estimate the Proportion (%) of the total area of channel within the Assessment Area that contains the following cover types:				
Woody Debris (e.g., logs, sticks, branches, and other natural woody material)	**Aquatic Vegetation** (e.g., grasses, rushes, sedges, water cress, water lily)	**Algae** (e.g., filamentous or floating algal mat)	**No Vegetation or Woody Debris** (e.g., water surface, dry bed)	**Total**
___________%	___________%	___________%	___________%	**100%**
Comments on Vegetated Condition:				

Figure 11. Vegetated Condition Assessment Form.

Importance of Site Information Collection

Conditions at a given site can contribute to the amounts and distribution of trash found in the general area. These conditions can change over time and for sites assessed regularly this information can also provide insight into any changes affecting trash at the site. Not all information is necessary, as some characteristics will not change, and some can be obtained through GIS exercises. For example, land use type in the general vicinity of the site can be estimated, the location of storm drains nearby to the site, and population size.

Other information collected shows little to no relationship to the amounts of trash present. For example, direct observations in the field indicate that the vegetation level found at a site may contribute to the retention of trash via entrapment. However, examining the relationship between the levels of vegetation and trash (Figure 12) we find that there is little relationship (R2 = 0.065).

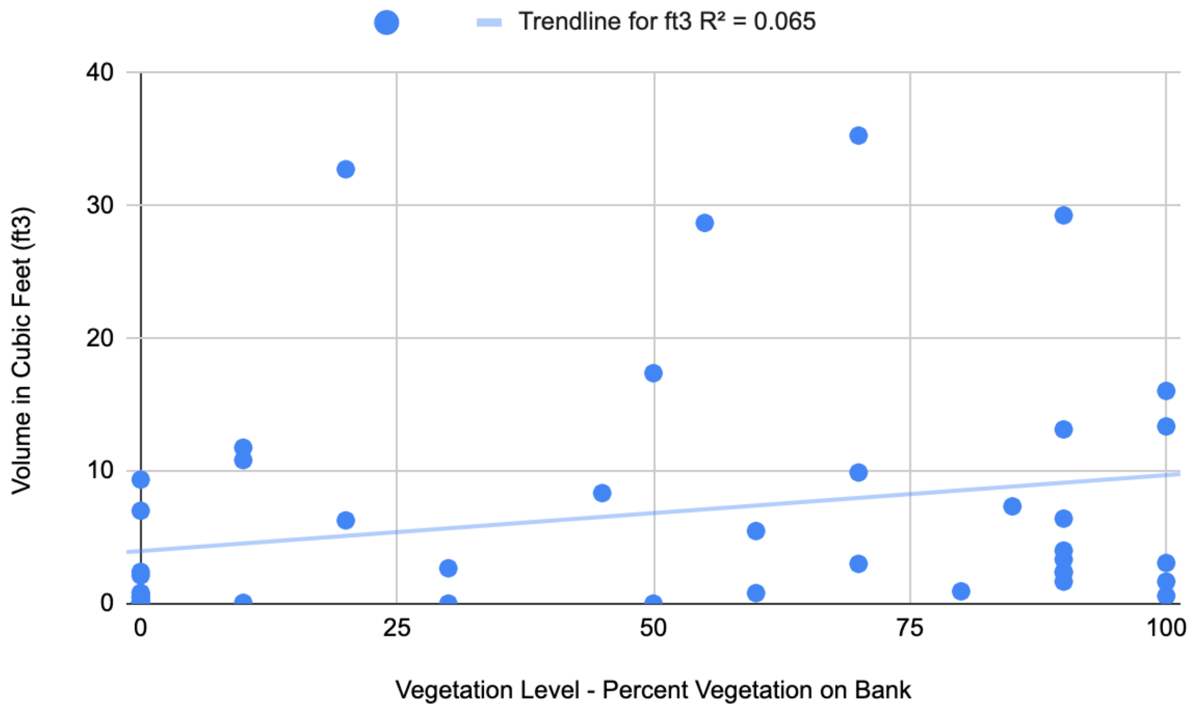


Figure 12. Vegetation Level versus the amount of trash in volume (ft3).

Site Information Recommendations

Collection of site information is important in determining characteristics within the river/stream that may influence the amounts of trash found at a site. Much of what is collected is easy to document and necessary for later analysis. Some of the information collected could be analyzed post sampling via GIS exercises back in the office. Measurements of the length and width of the area are important for normalizing the data and estimating the area covered by the assessment. For example, many of the concrete lined sites were wide and covered large assessment areas compared to the smaller earthen or natural sites. Computing a density of trash (i.e. items/meter^2) would normalize the amount of trash for comparison among sites and regions.

Consistency in Measurements

For measurements taken at assessment sites, it is imperative to ensure they are consistent among trash practitioners. The project team took part in the training exercises for both northern and southern California practitioners performing different methods. While the methods for estimating trash were different, the measurements taken to delineate the sites were similar. However, the Project Team noticed at least one difference in the measurements between the two region teams. For example, the measurement of Bankfull Width was different, with one team measuring the distance using the contour of the banks and stream, while the other team measured straight across. The Project Team recommends the former method be used, where the contour of the banks and stream are measured, as it leads to a more accurate measurement of area.

Keep it Simple

We recommend keeping the site information data collected simple. Photos of the assessment site and the general area are easy to collect and can provide some context when reviewing the data. Additionally, they can be used to train future crews when assessing the amounts of trash qualitatively.

For measurements taken at assessment sites, it is imperative to ensure they are consistent among trash practitioners.

Trash Condition Category and Site Score (Qualitative)

The qualitative assessment is a visual survey technique performed by at least two crew members (one being the Field Crew Supervisor) that subjectively documents the levels of trash within the survey area and estimates the relative contribution of trash. This method was originally part of the Rapid Trash Assessment developed by the Surface Water Ambient Monitoring Program (SWAMP) in California. The Bay Area Stormwater Management Agencies Association (BASMAA) revised the method and developed new categories based on descriptions as delineated in the field form (Figure 13). For this method, the Field Crew Supervisor first walked the entire assessment area and scored the site based on their "first impression" of the amount of trash observed. The trash condition was divided into four condition categories that include narrative descriptions of trash levels associated with a scoring range (1 – 12) as follows: Low (1-3), Moderate (4-6), High (7-9), Very High (10-12).

Station: ____________________ Sample Date: ____________________

Visual Assessment - Trash Condition and Pathways				
	Trash Condition Category			
	Low	Moderate	High	Very High
Description	• Effectively no or very little trash • On first glance, little or no trash is visible • Little or no trash is evident when streambed and stream banks are closely examined for litter and debris • One individual could easily remove all trash observed within 10 minues (100ft AA*) or 30 minutes (300ft AA)	• Predominantly free of trash except for a few littered areas • On first glance, trash is evident in low levels • After close inspection, small levels of trash are evident in stream bank and/or streambed. • On average, all trash could be removed by two individuals within 10 to 20 minutes (100ft AA) or 30 minutes to one hour (300ft AA) • Approximately 2-3 times more trash than the low condition category	• Predominantly littered except for a few clean areas • Trash is evident upon first glance in moderate levels along streambed and banks • Evidence of site being used by people: scattered cans, bottles, food wrappers, plastic bags etc. • On average, would take a more organized effort (more than 2 people, but less than 5) to remove all trash from the area. Removal of trash would take 10 to 30 mins (100ft AA) or 30 mins to 2 hours (300ft AA) • Approximately 2-6 times more trash than the moderate condition category	• Trash is continuously seen throughout the assessment area • Trash is distracts the eye on first glance • Substantial levels of litter and debris in streambed and banks • Evidence of site being used frequently by people (e.g., many cans, bottles, food wrappers, plastic bags, clothing; piles of garbage and debris) • On average, would take a large number of people (more than 5) during an organized effort to remove all trash from the area. Removal of all trash would take >40 minutes (100ft AA) or > 2 hours (300ft AA) • Approximately >2 times more trash than the high condition category
Site Score (30 meter)	1 2 3	4 5 6	7 8 9	10 11 12
Site Score (100 meter)	1 2 3	4 5 6	7 8 9	10 11 12

Figure 13. Trash Condition Categories and Scoring System form.

Observers physically walked on both banks and within or near the site (where feasible) to observe trash throughout the assessment area. Feasible conditions refer to flow conditions that allow the stream to be wade-able, in addition to conditions that would avoid impacts to migratory nesting birds and fish spawning. Trash that was visible outside of the assessment area was not included in the trash condition score but was noted in the comments section of the data form (BASMAA 2017).

Condition Method Repeatability

The comparison of assessments for this method show a high correlation (R^2= 0.928; Figure 14). In many cases the values assigned to one team were the same or similar to that assigned by the other team. The numbers in Figure 14 represent two teams that were both trained in assessing sites using this method. For example, one of the teams was the Southern California Stormwater Monitoring Coalition, who every year hold a training inter-calibration exercise as part of a larger survey effort. During this exercise, teams assess the same site and discuss any differences to ensure consistency when rating sites independently of one another. The Project Team, which was the other team, was also part of the inter-calibration exercise and acted as trainers. This can account for much of the similarity in measurements.

To ensure repeatability in this method, the Project Team recommends training be conducted prior to use of this method. This is particularly recommended for teams and individuals that have not conducted any previous trash assessments, as the concept of the highest amount of trash is difficult to conceptualize and visualize.

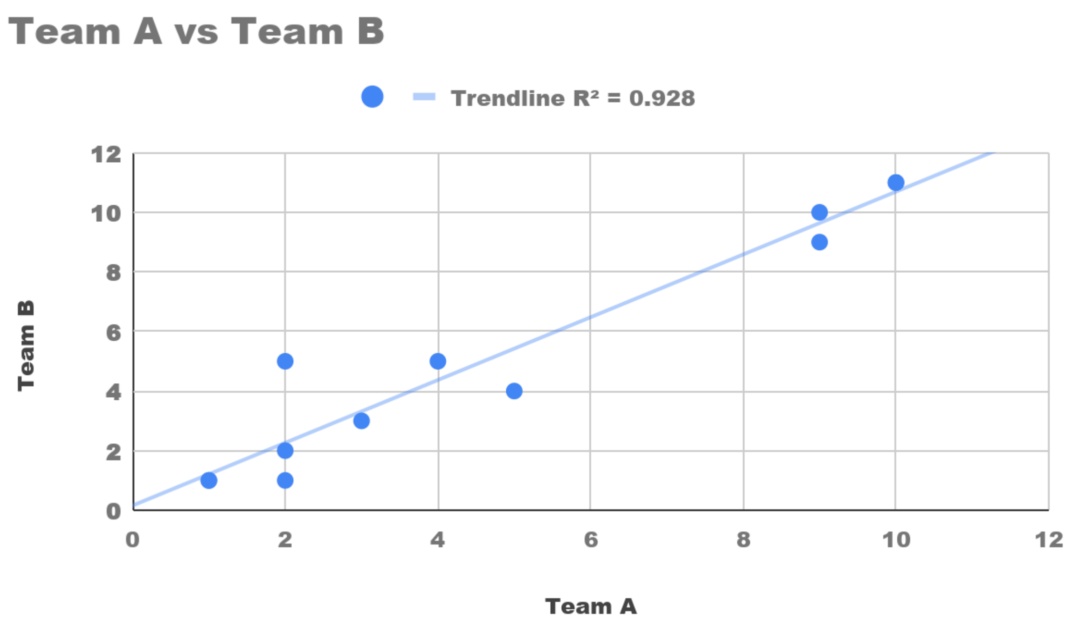


Figure 14. Qualitative scores (1-12) for two different teams assessing the same 30m sites. These are sites where two teams were assessing the sites separately. Each dot represents a specific assessment site.

Trash Condition Method Accuracy

Measurement of Trash Condition is difficult to assess for accuracy as this method is subjective and depends on the assessor's perception of trash conditions. The project team has found that consistency in this measurement is found when training is incorporated into this method prior to its performance in the field. One issue that has come up has been defining the "Very High" condition. In many cases, the limit on the highest amounts of trash that can be found has not been defined -- in other words, what constitutes a "12" on the scale. Different people have different definitions.

The BASMAA Trash Team attempted to help assessors in choosing a category by providing both descriptions of the levels of trash within each level and by providing pictures of the different levels of trash that would be defined within each category. The Project Team developed a mobile application (see Mobile Application section in this report) that used the descriptions developed by BASMAA but turned them into questions answered by the assessor with those answers leading to a recommended trash condition, based on the BASMAA rubric. Using this question-based technique compels a more deliberate process of evaluation, with answer-based evidence recorded as documentation. This is good approach to address potential bias, and would ensure evaluators to be more consistent.

Method accuracy was assessed for different assessment site lengths. The thought being that 30m may not be a representative enough length for a qualitative assessment (Figure 15). For most of the sites (67), the condition was assessed to be the same for both assessment site lengths. When sites were different, the 100m assessment sites were typically higher in qualitative scores (11), with a few being lower (5). This indicates that the 30m trash condition assessment in most cases, is representative of a given area. However, it is recommended that the 100m assessment be performed at sites that have characteristics (e.g., bridges, trails with public access) that might influence the site score.

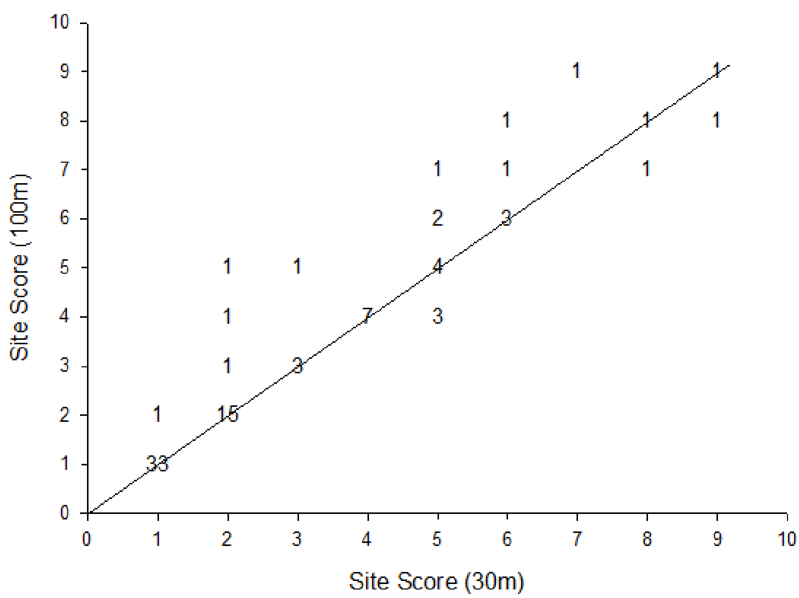


Figure 15. Condition method assessments for 30m versus 100m. The numbers in the plot indicate the number of sites with that score for both the 30m and 100m assessments.

Condition Method Conclusions and Recommendations

This method is the quickest and requires the least amount of resources; however, it is also the most subjective. The subjective nature of this method can decrease the repeatability of scores. Comparisons between trained teams show a high level of correlation and because of this, we recommend training for those using this method to increase consistency. Additionally, the use of a simple mobile tool can also be used to increase the consistency.

Tool Use to Increase Consistency

One way the Project Team approached the subjectiveness of this method is through developing a mobile application (see section on Survey Application Development below). Using the mobile application, the assessor was asked a series of questions based on the Trash Condition Categories (see Figure 13), and once they responded an algorithm was used to suggest what the score should be. The assessor could accept that score or change it based on other criteria considered.

Maximize the Area Assessed

For this method, the Project Team assessed both 30m and 100m sites. While in most cases the site score was the same for both assessment lengths, in some cases the site condition score was higher in larger areas (see Figure 15). Since this method requires few resources and time, we recommend maximizing the assessment area to obtain the most characteristic estimate of the site condition.

Trash Volume Measurements (Quantitative)

The quantitative volume assessment is a survey technique developed by BASMAA. This method includes estimating the total volume of trash at a site and providing rough estimates of volume by pathway. The pathways included are Litter/Wind, Illegal Encampment, Illegal Dumping and Unknown (stormwater or unknown upstream sources).

Quantitative volume measurements were performed at sites in this study by collecting all trash from the assessment area. In Northern California, this was done either by the BASMAA field crew, performing part of their scheduled surveys, or by the SFEI team if it was not a BASMAA site. In Southern California all volume surveys were conducted by the SCCWRP team. Trash items that were not completely visible during the assessment and/or could not be safely accessed by field crew were not included in the assessment but were noted in the comment section on the data collection form. Partially visible trash included items on the bottom of the wetted channel. Inaccessible trash included items trapped in tree branches, dense vegetation (e.g., blackberry bushes) or on steep banks that could not be safely accessed.

Estimate Trash Volume

After completing a qualitative assessment, all trash was collected from the assessment area and its volume was estimated. Trash outside of the defined assessment area was not collected or quantified as part of this protocol. Trash in the first season of Northern California volume method assessments was assessed for trash pathway based on definitions assigned by the BASMAA Trash Monitoring Program Plan (Table 2). Pathways included Litter/Wind, Illegal Encampment, Illegal Dumping and Unknown (stormwater or unknown upstream sources). Each item was assessed for its potential trash pathway and placed into a bucket or area representing the trash pathway and the estimated volume for trash in each category was recorded (Figure 16).

Table 2. Trash items typically associated with four types of transport pathways (from the BASMAA Trash Monitoring program plan).

Trash Pathway	Trash Characteristics	Potential Location in Assessment Area	Example Trash Items
Litter/Wind	- Light weight - Distributed evenly, recent/not worn	- Adjacent to or under freeways and road crossings - Near roadways, bike or foot paths adjacent to the water body	- Fast food items - Paint spray cans - Carryout plastic grocery bags - Paper - Styrofoam

Illegal Encampments	- Large items - Dense, multiple piles near current or abandoned camping site - No sign of water damage	- Adjacent to camps or trails - Banks, above and below high water mark - Under bridges	- Mattresses - Fast food items - Bagged trash - Large items - Fabric and cloth - Cardboard/paper - Metal cans/debris - Glass Bottles/pieces - Food Containers
Illegal Dumping	- Large items - Recent - Large piles, adjacent to roads	- Directly upstream or downstream of bridges - Near roadways	- Furniture - Bags of trash - Construction debris - Fabric and cloth - Mattresses - Tires
Unknown/Other (e.g., Stormwater, and Unidentifiable Upstream Sources)	- Small, persistent, transportable - Old, worn, water damaged - Integrated with vegetation, debris - Well distributed and mixed with debris	- Wetted channel - Banks below high water line - Directly below outfalls	- Polystyrene food ware - Cigarette butts & wrappers - Food wrappers - Plastic bottles/cups - Plastic straws/caps - Carryout plastic grocery bags - Rubber balls/tape - Paper fragments

Relatively small trash items associated with each category were collected in 5-gallon and 2-gallon buckets or super heavy-duty trash bags of a known size. The outside of buckets were marked with a permanent marker in 0.5-gallon increments. Once the bucket was full (i.e., level with the top of the bucket), it was emptied into a super heavy-duty plastic garbage bag (e.g., 30 gallons). Trash in partially filled buckets was transferred to 2-gallon buckets, and volume was estimated using 0.5-gallon increments. For trash volumes less than 0.5 gallons, "< 0.5" was marked on the field data collection form. Small trash items that were included in buckets/bags include the following: food wrappers, spray paint cans, paper products, glass and plastic bottles, takeout food containers and utensils, clothing/shoes, sports balls, spray paint cans, small styrofoam, aluminum, steel and tin cans, cigarette butts, single use plastic bags, paper products, cardboard, small automotive related items.

Estimated Volume of Trash Removed

▶ **Record total volume of trash associated with each trash pathway that was collected in the assessment area. For small items collected in buckets or bags, use 0.5 gal increments. For large (unbagged) items, use 0.5 ft³ or yd³ increments.**

Trash Pathway	Trash Volume (Un-compacted)					
	Small (bagged) Items[1]				Large (Unbagged) Items[2]	
	# Buckets	Bucket Size (gal)	# Bags	Bag Size (gal)	Volume	Unit (circle)
Litter/Wind						
Illegal Encampment						ft³ yd³
Illegal Dumping						ft³ yd³
Unknown (e.g., Stormwater, or Unknown Upstream Sources)						
Total Trash (Sum of the above rows)						ft³ yd³

[1] Small items may include: Food Wrappers, Takeout Food Containers and Utensils, Glass and Plastic Bottles, Clothing/Shoes, Sports Balls, Spray Paint Cans, Small Styrofoam Aluminum, Steel, and Tin Cans, Cigarette Butts, Single Use Plastic Bags, Small Automotive Related Items, Biohazards (Syringes, Diapers, Human Waste, Pet Waste), Paper Products, Cardboard

[2] Shopping Carts, Mattresses, Coolers, Furniture, Appliances, Tires, Bicycles, Construction Debris, Automobile Parts and Large Bags of Trash

Figure 16. Estimated Volume of Trash Removed Form From the BASMAA Trash Monitoring Program Plan.

Trash was placed in buckets and bags and was not compacted. Garbage bags were not filled with more than 40 to 50 pounds of material. If material contained sharps or large objects, the material was "double bagged", as necessary. Multiple garbage bags were used per assessment site, as needed. The total number buckets and volume of collected trash was recorded on the Estimated Volume of Trash Removed Form. All biohazards and hazard waste were separated and dealt with appropriately by the site leader.

Material that was too large to be placed in buckets or bags was estimated for volume visually. Estimates of large items (e.g., construction materials or appliances) were made in cubic feet or cubic yards and recorded on the Estimated Volume of Trash Removed Form. Large items include, but were not limited to, the following: shopping carts, mattresses, coolers, furniture, appliances, tires, bicycles, construction debris, automobile parts, and large bags of trash.

A total volume was calculated for all items, small and large, and recorded on the data sheet.

Volume Method Repeatability and Method Accuracy

Repeatability for the Volume Method was difficult as sites where two teams assessed the trash, were often visited on different days and only one team performed this assessment. Additionally, comparisons were confounded by the collection of the trash during the performance of this method. However, comparison of results for both the Tally Method and the Qualitative Site Condition Score method give insight into the repeatability of this method. As we saw with the Tally Method (described in a later section), particularly with the extinction curves, most trash is removed after the first 1-2 passes at a site. The sites requiring more passes, typically had smaller trash items that were missed, i.e. small plastic pieces. This indicates that larger, more volumetric trash items are picked up and only smaller trash items may remain; hence, not making huge differences in volume measurements.

Volume Method versus Qualitative Method Site Condition Scores were highly related (R^2 = 0.894) indicating that site condition is a good surrogate for measuring volume. Site condition was found to be very repeatable with numbers not often different and only varying by 1-2 when they did vary.

Additionally, there was a pattern of high correlation between volume predictions using UAS methods and directly measured volume totals, which supports efforts to explore volume prediction methods.

Determining the upstream source of trash proved to be difficult when classifying based on trash type and most trash was placed in the Unknown category. Repeatability in this case was difficult because of lack of practitioner consensus in what fell into the Litter/Wind, Illegal Encampment, Illegal Dumping, and Unknown categories.

Volume Method Resources

Resources for this method are considered high, particularly for labor. Times varied for this method and ranged from 20 minutes to 3 hours and 30 minutes for a 300 ft site. The average amount of time it took to perform this method was 1 hour and 31 minutes. This, in many cases, was less time than it took for the Tally Method, with few exceptions. Additionally, equipment in the form of buckets, bags, trash pickers, and gloves are required for both measuring the trash and ensuring the safety of trash assessors.

Volume Method Conclusions and Recommendations

This method measures the volume of trash in a given assessment area, categorized into observed pathways. The information provided indicates the amount of space anthropogenic trash takes up in the area and does not represent the types of trash found (i.e. plastic, metal, glass, etc.). The estimates of upstream sources used in this study were difficult to assess. We recommend the following to those using this method:

Consider Using Site Condition Score as Alternative Method

Site Condition Score is highly related to the amount of trash as measured by volume, therefore, where applicable we recommend the use of the Site Condition Score in place of the Volume Method, particularly when resources are low, or trash extraction is otherwise not feasible. This surrogate was also recommended by BASMAA in its report San Francisco Bay Area Receiving Water Trash Monitoring: Pilot-Testing of Qualitative and Quantitative Monitoring and Assessment Protocols (BASMAA 2020) as they saw a similar correlation between Volume measurements and Site Condition Scores. Those considering the use of the Site Condition Scores should take into account the need for training trash practitioners in the method to ensure consistency in this subjective method.

Simplify Trash Source Categories

We recommend not estimating the pathways suggested for this method of Litter/Wind, Illegal Encampments, Illegal Dumping and Unknown as these categories are difficult to assess. Instead we recommend breaking the trash into material categories that relate to the Tally Survey. These categories are based on the composition of the trash item and include Plastic, Miscellaneous, Fabric & Cloth, Biodegradable, Biohazard, Construction, Glass, Large, and Metal. Breaking the trash into these categories allows for the identification of categories making up the majority of the volume and provides information on those that should be targeted for potential source control.

Trash Count Method (Quantitative)

This method was originally developed as part of the Surface Water Ambient Monitoring Programs (SWAMP) rapid trash assessment developed in 2002. The Southern California Stormwater Monitoring Coalition (SMC) has modified this method and added it to their annual Regional Stream Surveys in 2011-2013 and 2018-2019. As part of their 2018 Regional Stream Survey, the SMC sampled sites during the summer throughout southern California. At a subset of their sites, the Field Testing Team for this project followed them and assessed the same sites. In addition, many other agencies, including non-profits and regulated organizations, assisted with the testing.

For this method, trash was quantified by recording the specific types of material and their quantities. Trash was divided into nine major categories, which include Plastic, Miscellaneous, Fabric & Cloth, Biodegradable, Biohazard, Construction, Glass, Large, and Metal (Figure 17). Under each of these categories trash was broken out into more specific items. For example, for the Plastic trash category, items such as bags (both reusable and single), cigarette butts, foam food containers, and hard plastic pieces are listed and can be tallied. In some cases, specific items (Table 3) that were hard to count in large numbers, were estimated and binned into two categories: Moderate (M) = 11-100 pieces and High (H) >100 pieces.

Table 3. Items That May Be Estimated on the Trash Tally Form.

Bag pieces	Foam pieces	Yard waste/leaf piles	Aluminum foil pieces
Soft plastic pieces	Wrapper/wrapper pieces	Glass pieces	

StationID:__________________ Date:__________________ Initials:________________

Plastic	Tally Marks	Total
Bag - reusable		
Bag - single use		
Bag Pieces*		
Balloons		
Beverage Bottles		
Chip Bags		
Cigar Tips		
Cigarette Butts		
Container Cap/Pieces		
Cups		
Foam Balls		
Foam Cups		
Foam Food Containers		
Foam Other Containers		
Foam Packing Peanuts*		
Foam Pieces/Pellets*		
Foam Plate		
Hard Plastic Container		
Hard Plastic Pieces		
Lid		
Lighters		
Pens/Markers		
Plastic 6-Pack Holder		
Plastic Bottles		
Plastic Pipe		
Plastic Straw Wrapper		
Plates		
Single Use Container		
Soft Plastic Pieces*		
Straw/Stirrer		
Tarp		
Tobacco Wrapper/Pieces		
Wrapper/Wrapper Pieces*		
Plastic Other		
Fabric and Cloth	Tally Marks	Total
Natural (Cotton, Wool)		
Shoes		
Synthetic Fabric		
Tent/Sleeping bag		
Fabric Other		
Large	Tally Marks	Total
Furniture/Appliances		
Garbage Bags of Trash		
Shopping Carts		
Tires		
Large Other		
	Total Items:	

Biodegradable	Tally Marks	Total
Food Waste		
Paper/ cardboard		
Yard Waste/Leaf piles*		
Biodegradable Other		
Biohazard	Tally Marks	Total
Condoms		
Dead Animals		
Human Waste/Diapers/TP		
Medical waste		
Pet Waste		
Biohazard Other		
Construction	Tally Marks	Total
Bricks		
Concrete/Asphalt		
Fabricated Wood		
Rebar		
Construction Other		
Glass	Tally Marks	Total
Glass Bottles		
Glass Pieces*		
Glass Other		
Metal	Tally Marks	Total
Aluminum Foil pieces*		
Aluminum or Steel Cans		
Auto Parts		
Metal Bottle Caps		
Metal Pipe/Bar Segments		
Nails, Screws, Bolts, etc.		
Small Batteries		
Spray Paint Cans		
Wire (barb, chicken, etc.)		
Metal Other		
Miscellaneous	Tally Marks	Total
Ceramic Pots/Shards		
E-waste		
Foam rubber		
Hose/Hose Pieces		
Mylar balloons		
Rubber/Rubber Pieces		
Sports Balls		
Waxed Paper Cups/Plates		
Misc. Other		
	Total Items:	
	GRAND TOTAL:	

*** These items may be binned if abundance is greater than 10 pieces as follows:**
M = 11-100 pieces
H = > 101 pieces

Figure 17. Trash Tally form.

Assessment areas varied from 100 feet to 300 feet. The width of the assessment area was delineated from the thalweg to bankfull width on the left and right bank (face downstream to determine left or right bank). The team began their survey on either the right or left bank, walking slowly while visually scanning for trash. The team scanned an area within a shoulder width zone in heavily vegetated sites to avoid missing small or partially covered items. A larger scan width was used at sites with little or no vegetation. The systematic scan approach was used while walking the assessment area.

Trash Count Method Repeatability

Method repeatability is important to determine if measurements both within and among groups are comparable. For this exercise multiple teams assessed the same site and identified and tallied items. Three teams assessed an earthen/concrete lined site within San Juan Creek, located in Orange County, California. The teams consisted of 3 people each and were directed to walk in a systematic pattern. One team was the project team (Team A) and did the counts both in the field and in the lab after the trash was collected. The other two teams were trained and performed the assessment on the same day.

The amount of trash collected for each category was relatively similar for categories containing small amounts of trash (Figure 18). Larger differences were found for the Plastic category. Two teams found similar numbers of plastic (189 vs 174) and one found a much lower amount (130). The project team removed the items from the site, counted them in the lab and for plastic found a higher amount (232) than that identified in the field. The larger amount counted in the lab is likely attributed to the breaking of plastic into smaller pieces during collection and transfer. Much of the plastic in the field is often brittle from photodegradation and will break easier. Therefore, using the amount of plastic Team A found in the field as a more reliable estimate of the total, Team B identified 68% of the total, and Team C 92% of the total.

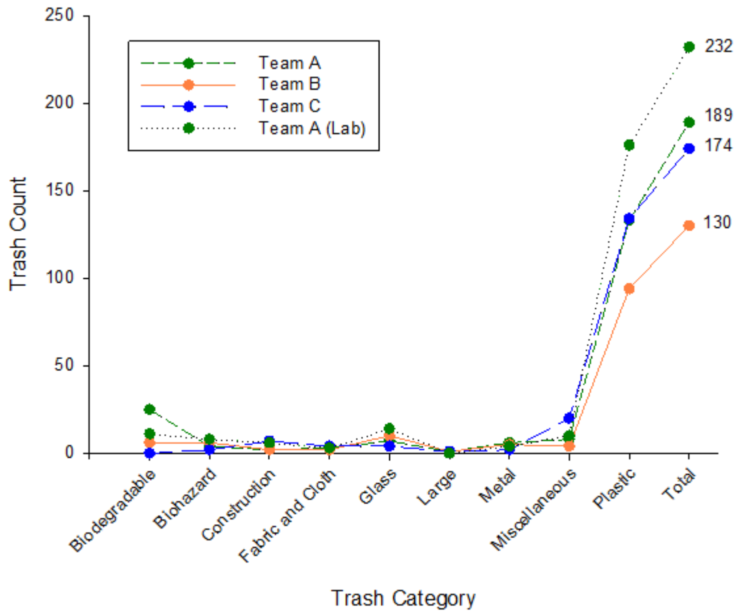


Figure 18. Comparison of the amounts of trash within each category as identified and counted by different teams. Team A was the project team.

Examining the amounts of plastic of specific types, differences are seen between the teams. Cigarette butts were identified in similar numbers by both Teams A and C, but Team B did not identify many at all. The overall differences in the plastic category on a whole are also reflected in the differences in numbers for hard and soft plastic pieces. Pieces are often small and harder to see in the field, perhaps accounting for some of the differences. Other differences are likely attributable to search image patterns of the groups/individuals and what they may have an easier time seeing. For instance, while Team B was lower on counts in almost every category, they were higher in Foam Balls and in Wrapper/Wrapper pieces (compared to Team C).

Many of the differences can be attributed to differences in definitions of the debris items themselves. Conversations with participants in this exercise, indicated what one person might call something, another would have a reason for calling it something different. For example, a degraded wrapper piece might be identified as a soft plastic piece or vice versa. It is recommended that for this method, the definition of items should be clear, and participants should be minimally trained to identify them.

It is recommended that for this method, the definition of items should be clear, and participants should be minimally trained to identify them.

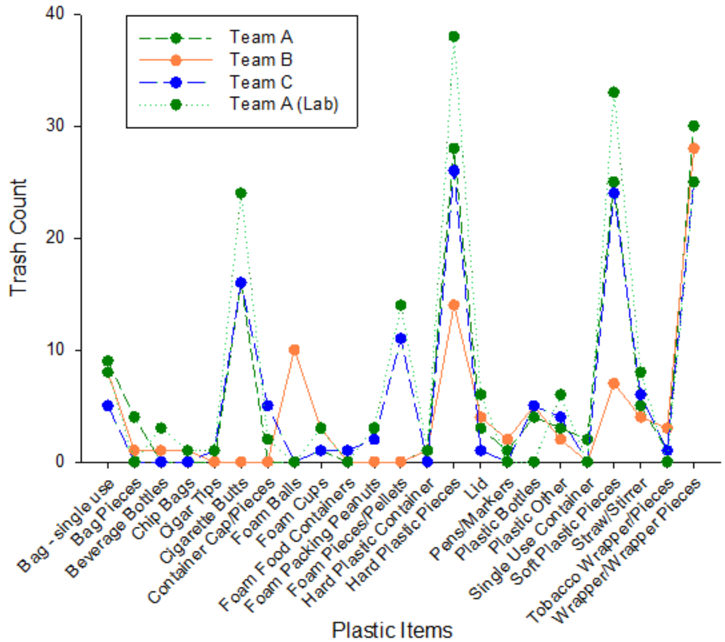


Figure 19. Comparison of the amounts of specific plastic items by team. Team A was the project team.

A second site was assessed by multiple teams in a large concrete-lined channel in the Los Angeles River. Two teams identified and counted trash and again the main differences were seen in the plastic category (Figure 19). Team B counted 248 plastic items, more than a two-fold difference from Team A who measured 101 plastic items. When examining the items, the teams differed the most in, five items were identified and counted more often by Team B, the highest two being Foam Packing Pieces and Wrappers/Wrapper Pieces. Team A identified and counted more Beverage Bottles than Team B. In this case, identification did not appear to be an issue.

This site, while completely concrete, was mostly wet with flowing water. One team was more cautious about getting into areas that were wet, due to the lack of proper gear, suggesting that teams assessing sites should 1) perform reconnaissance prior to site assessments and 2) assure that proper gear is available for the assessment.

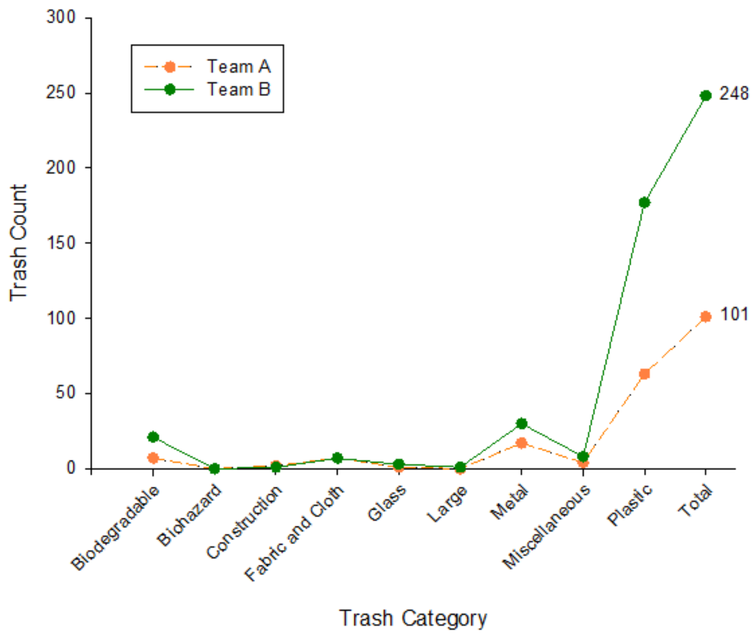


Figure 20. Second team assessment comparison of trash category counts.

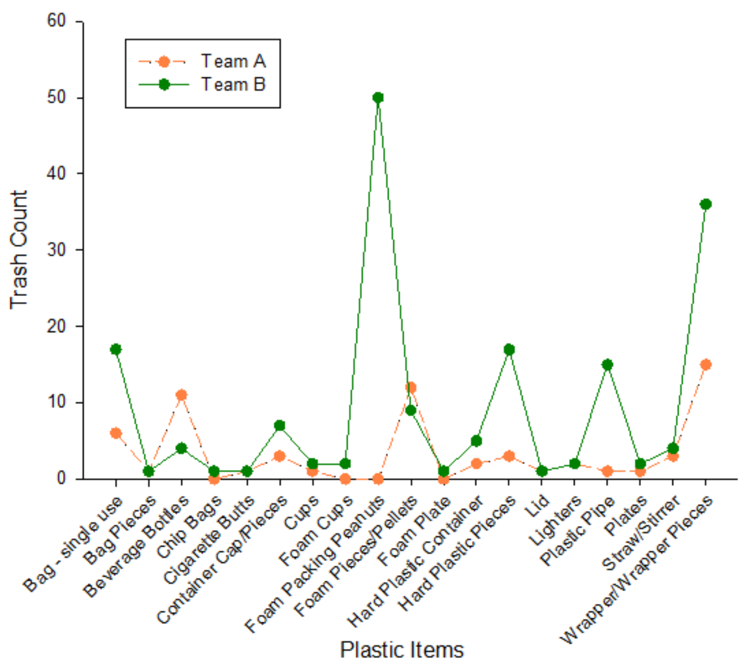


Figure 21. Second Team Comparison of counts of specific plastic items.

Finally, comparisons were made between different types of teams and the project team (Figure 21). We included both trained (SMC and NGOs) and untrained groups in this assessment. The SMC (Southern California Stormwater Monitoring Coalition) group as a whole, consisted of highly trained staff, most of whom had been performing trash assessments as part of a 2011-2013 study. The NGOs were trained, but performed assessments less frequently, and the newly assembled Training Groups were participating in the assessments as part of their training. While the data show the SMC trash counts were close to the project teams counts, the SMC-performed trash assessments showed lower amounts of trash (without the one outlier). The Training Groups and the NGO's counts varied much more than the Project Team; however, their assessments had much more trash, which may have accounted for more of the variation. In general, we found the groups with more training were closer to the Project Teams numbers, indicating a level of repeatability for this method

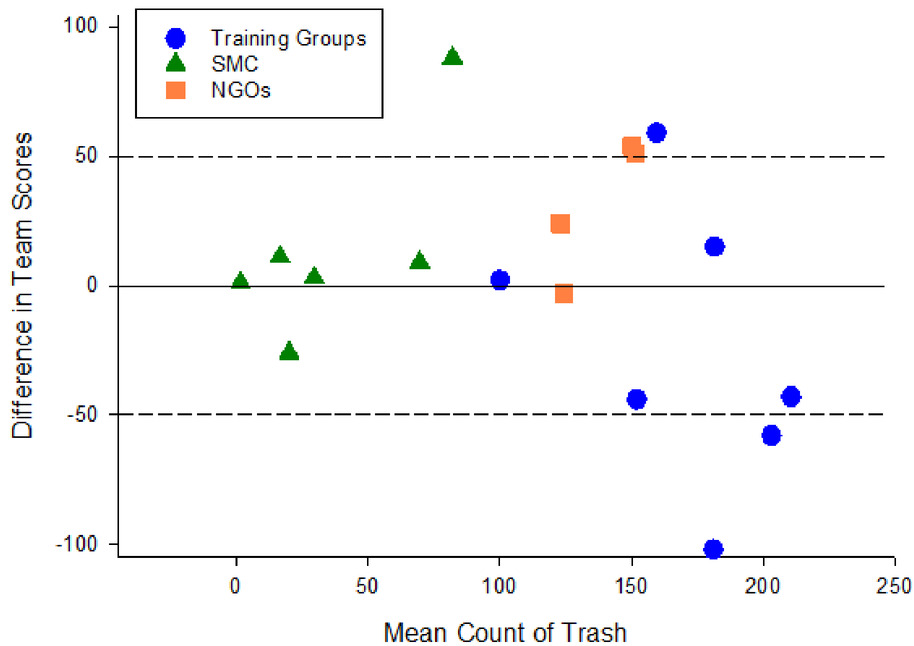


Figure 22. Differences in Team Scores which reflect the different in counts of trash at an assessment site versus that of the Project Team.

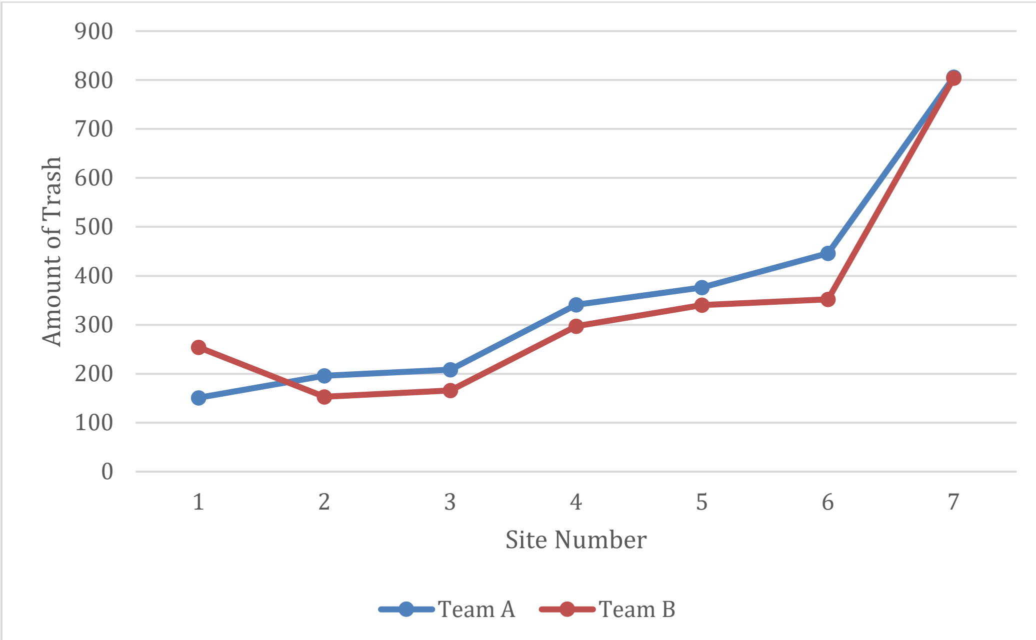


Figure 23. Team differences for trained teams in Bay Area sites.

Trash Count Method Accuracy

To determine method accuracy for the Tally Method, an extinction curve was performed. This involved using different teams of people to systematically search an assessment area for trash, making several passes until no trash was found. In most cases, the majority of the trash was removed during the first pass, and after the second pass the amount of trash was near zero (Figure 24). This extinction curve exercise also pointed out some differences in trash size. For example, Site 1, with the highest amount of trash, took the most passes (5) to get to zero trash; however, Site 1 had a very small volume (Figure 25) and most of the trash was identified as small plastic pieces, indicating the smaller the trash at a given site the more trash might be missed.

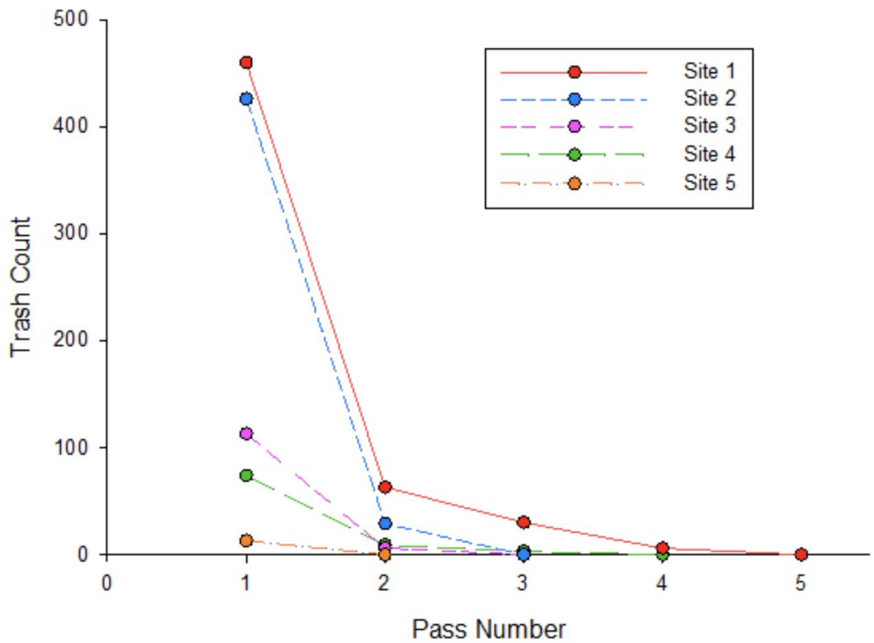


Figure 24. Extinction curves showing the amount of trash removed during multiple passes at an assessment site (lines and colors represent different sites).

Figure 25. Site 1 (from figure 24) trash showing a small volume despite having a large amount of trash.

Trash Count Method Resources

The amount of resources needed to perform this method varies based on site condition and whether the trash would be picked up. Material resources are minimal if the trash is not to be extracted and include data sheets, writing implements, clip boards, and measuring tape(s). If the trash will be collected, then trash pickers, buckets/plastic bags, and gloves should be required. Staff resources for this method should be the most heavily considered. Typically, two 30m assessment sites per day can be done by 2-4 people depending on travel time to a site and the amount of trash at a site. Higher level trash sites can take approximately 2+ hours to assess a 30m site for this method (Table 4). Much more time is needed for 90m assessment sites with higher levels of trash.

Table 4. Time estimates for both 30m and 60m assessment sites. These time estimates represent targeted high trash sites in southern California. (Values in gray represent 30m assessment sites).

Station	Date	Number of People	Distance Assessed (m)	Trash Amount[2]	Time (hh:mm)
San Diego Creek	9/13/2019	2	30	270	1:49
San Diego Creek	9/13/2019	2	90	574	3:20
Trabuco Creek	9/24/2019	2	30	118	0:31
Trabuco Creek	9/24/2019	2	90	432	1:34
Aliso Creek	9/25/2019	5	30	49	0:21
Aliso Creek	9/25/2019	5	90	288	1:08
Greenville Banning Channel	10/18/2019	4	30	730	2:29
Greenville Banning Channel	10/18/2019	4	90	1412	3:43
Los Angeles River	10/23/2019	4	30	1415	2:11
Whittier Narrows	9/20/2019	6	30	678	1:58
Whittier Narrows	12/11/2019	6	30	978	2:00

Conclusions and Recommendations

Training increased both the accuracy and repeatability of this method. Many of the differences can be attributed to differences in definitions of the trash items themselves. We recommend developing a robust trash library with images, definitions, and examples to convey the different trash types measured in this survey. Conversations with participants in this exercise, indicated trash in one category was often placed in different categories by other persons. In other words, what one person might call something,

[2] Conservative estimates – some trash was estimated, and the lower values were used in the totals (for example Small Plastic Pieces might have fallen in the H(High) >100 pieces category and were given a total of 101 used to calculate the total trash amount).

another would have a reason for calling it something different. We recommend, in addition to the robust library, that participants should be furnished with some training in identifying trash. This can easily be done through a field exercise but can also be accomplished by having would-be participants watch a short video describing the library with examples. Differences in numbers between different groups is expected but variation decreases when training is provided.

Survey Application Development

To aid with data collection, the software engineering team at SFEI leveraged an off-the-shelf survey tool, Esri's Survey 1-2-3, to standardize the fieldwork data. In practice, we encountered positive and negative outcomes associated with the use of this survey tool, which might be generalized more broadly to other digital-input methods, to varying degrees.

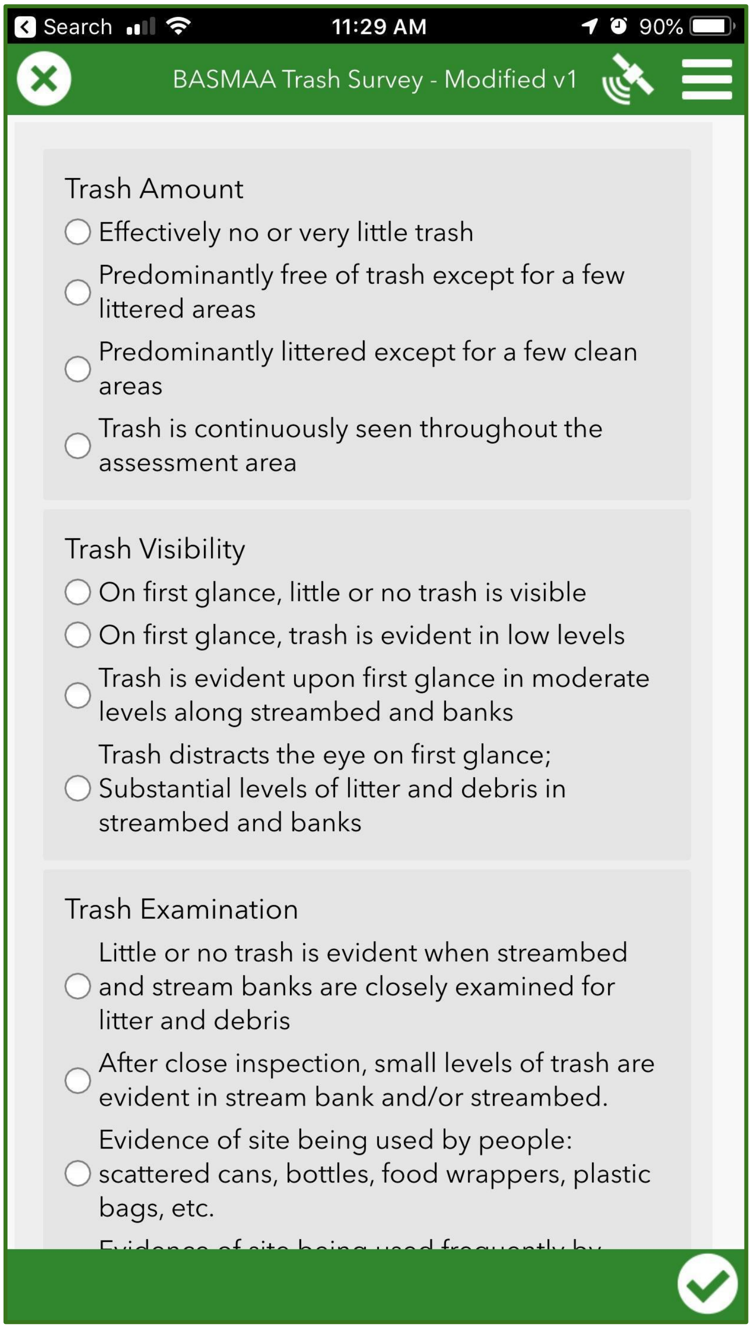


Figure 26. Screenshot of Survey 1-2-3 showing the digital interface for BASMAA trash surveys.

The primary goal for the use of this digital tool was to eliminate the use of paper, while simplifying data collection processes. Among the key benefits of the tool were:

1. Improved data validation.

   For key fields, the tool would only accept valid values, whether alphanumeric or selected from a list. Furthermore, in many cases, the methods call for relative percentages associated with the site characterization. The automatic validation of the percentage allocations -- ensuring that all percentages sum to “100” -- reduces the likelihood of practitioner error in the field.

2. Direct entry into spreadsheet formats (semi-structured data).

   The survey tool exports into a comma-separated values (CSV) format. This is a standard and simple format that can be analyzed in a host of tools. This reduces the work associated with data entry, and also reduces the incidence of typographical errors producing inaccurate data.

3. Automatic collection of location-based data.

   For site characterization, precise location data is important for the general and specifics of the method. The Esri tool facilitates this collection with a simple click.

4. Images and guidance can be embedded close to the prompts for timely instruction.

   Each method comes with a guide. The materials associated with each method can be parsed and featured in the data collection tool, ideally promoting consistency as the reference material is always close at hand, and key visual guides can ensure greater consistency among practitioners.

5. Easy collection of imagery and automatic association of imagery with the site.

   The tested method uses imagery as verification. The tool collects the imagery and keeps the images connected to the related sites via links.

6. No active connection is required.

   The tool can be pre-loaded with the survey and all of the needed resources to conduct the survey in the field, even when an active internet connection is not available in the field.

We also encountered several constraints and limitations associated with the tool. These included the following:

1. Default data schema is limited/confining.

   Survey 1-2-3’s default schema required a high degree of customization. And even with these interventions, the results were not without shortcomings. For instance, the values needed to facilitate automatic calculation were awkward and cannot be excluded from the data export.

2. At present, Survey 1-2-3 cannot collect polygonal data.

   The tested methods implicitly or explicitly collect site characteristics that form spatial polygons. Esri’s tool cannot accommodate the drawing of polygons, which would otherwise be a boon to efficient data entry.

3. The tally methods are most efficient using paper.

The team found that the tally method, which is already labor intensive, slowed to a crawl when individually describing individual trash items on a digital interface. Paper, even with its inherent limitations, exceeded the performance of any survey we could deploy using Survey 1-2-3 for tally-based surveys.

Bearing in mind these advantages and disadvantages, among the most important strides effected through the digital survey was the use of the BASMAA matrix to improve the support for qualitative characterization of trash levels in the landscape. The BASMAA method carefully details the individual characteristics associated with the different qualitative scores. As shown in Figure 20, the survey solicits feedback on the individual characteristics that might otherwise be overlooked in the rapid practice of the condition assessment. The survey then calculates the suggested trash level based on a mean score associated with the different responses. The practitioner can select her own score based on professional judgment and experience. However, the suggested score, based on her individual responses to the matrix prompts, is also recorded. This means that analysts can subsequently determine the degree of discrepancy among the scores suggested by the system and the overall practitioner-defined values. The results of this analysis might be an occasion for practitioner inter-calibration or additional training. Essentially, this small innovation offers an opportunity for additional field-level quality control.

Unoccupied Aerial System-Based Observations

For 20 site assessments, we augmented the conventional methods with UAS-based surveys. The survey team deployed a 3DR Solo, equipped with a Sony R10C sensor and later a 3DR Mavic 2 Pro, equipped with a Hasselblad L1D-20c sensor. Both UAS are able to produce comparable imagery quality. For flight planning and imagery post-processing, the team used SiteScan software, also engineered by 3DR. SiteScan offers data hosting to facilitate rapid and easy sharing of the resulting imagery.

Timing and Coordination

To maximize the comparability of the results gathered by on-site assessments with those derived from the UAS-based imagery, the timing of the flights is critical. Trash is ephemeral in its location -- mobilized by wind, water, and other factors -- so the flights were conducted as close to the time of trash extraction as possible.

The survey teams coordinated their flight planning to coincide with on-site surveys. In cases when weather or other circumstances prohibit flights on the same day as the conventional assessments, the UAS team conducted their flight on the day prior to the in-stream assessment.

Personnel and Safety Measures

Every flight was planned, approved, and conducted by an Unoccupied Aerial Vehicle (UAV) pilot, licensed by the Federal Aviation Administration (FAA). Although not required by FAA regulations, the UAS team included a spotter who helped to ensure the safety of the vehicle and site, with respect to any visible obstacles.

The vehicle is subject to a broad set of regulations (https://www.faa.gov/uas/media/Part_107_Summary.pdf). Because the associated technology is evolving quickly, the regulations seek to keep pace, with revisions published on a regular basis. The team observed all formal regulations.

We recommend adhering to the practices outlined below.

Field Preparation

Prior to going into the field, the team undertook the following preparations:

- Collect and review relevant background information, including: prior site photos, if any, and GIS maps, if any.
- Receive permission for access and survey from landowner of study area
- Notify local law enforcement to activities. This is not required but considered good practice.
- Check for any wildlife considerations in the area and adjust flight plan accordingly (elevation, timing etc.).
- Identify the equipment needs for field observation. (If protocols are further refined in the field, document the revisions)
- Coordinate equipment and materials, including: UAV, GPS, GCP markers, ground station, iPad, camera/sensor, GPS, extra batteries, map with aerial imagery, flight plan, watch, anemometer, handheld digital camera, photo point monitoring log, a notebook, orange vests to provide visibility, safety gear (e.g., fire extinguisher, water, sunscreen, hat, food, first aid kit, etc.).
- Maintain and calibrate field equipment, as appropriate. Ensure that UAV is operating nominally. Bring extra propellers and batteries.
- Ensure that UAV is registered with Certificate Number visible on the UAV
- Ensure that in addition to FAA licensed UAS commercial remote pilot that there is also a visual observer during each flight.
- Provide Flight Plan ahead of flight to interested partners. Flight Plan to include, flight area, altitudes, and launch locations.
- Ensure UAS is covered for liability
    - SFEI has insurance for the vehicle through Allianz Global Corporate and Speciality ($1M limit liability)
    - SFEI also has general insurance
- Check FAA regulated airspace and any new restriction to ensure compliance.
- Check weather compliance day of flight. Ensure 500ft above maximum altitude and cloud ceiling and 200ft horizontal from any low cloud banks not allowing for 500-ft clearance.

In-Field and Flight Conduct

Once in the field, the following materials were available on request and the team adhered to the following conduct:

- Pilot in Command possessed a copy of FAA issued UAS Remote Pilot License.
- Visual Observer and Pilot were present for each flight. Both staff wore highly visible fluorescent vests to provide visibility.
- UAS flight crew were able to provide flight plan, UAV Certificate Number, and any permission obtained for flight authorization.
- Flight crew followed all FAA rules. Some regulations of note:
    - Do not fly over 400ft (target of 100ft elevation for survey flights).

- Fly between 30 min prior to sunrise and 30 min after sunset; however, optimal flight time is close to noon in order to capture best lighting for post processing and analysis. Avoid flying at solar noon over water when capturing nadir imagery.
    - Maintain visual line of site with aircraft at all times by remote pilot in command or visual observer.
    - Do not fly UAV directly over people not part of active flight.
- UAV was flown using mission planner with the ability to assume manual control the vehicle.
- Alerts were provided to pilot at 25% and 15% battery, returning home at 10% capacity or if the voltage drops below 14 V, whichever happens first in order to ensure safe landing of the vehicle.
- Wind speeds were taken into account to ensure safe, controlled flight and to ensure the flight could be completed with existing batteries.
- All ground control markers were retrieved after fights in the area were completed.

Imagery

The vehicle flew a path designed to overlap with the targeted assessment area, taking high-quality, still images as it flew. Over the course of the flight, the images sometimes numbered in the hundreds.

Flying at 30m altitude, the team was able to produce imagery with resolution of approximately 1cm per pixel. Imagery of this resolution facilitated identification of most trash items in the environment.

Figure 27 illustrates the results gathered from a test flight, whereby the sensor fires at regular intervals to capture individual, high-resolution images. (While it is possible to use video as a source of the images, the resolution is reduced and thus less suitable for our purposes.)

Figure 28 shows the results of the post-processing of the individual images into a single orthorectified orthomosaic. This single image can serve as the primary source of further analysis. However, if the team encounters any visual artifacts or distortions, they may consult the original, individual image associated with the suspect location.

Figure 27. Trash survey flight over the Wildcat Creek, showing an orthomosaic of the site, each bubble indicating a separate picture, December 7th, 2018

Figure 28. Test survey flight over the Wildcat Creek, showing an orthomosaic of the site, December 7th, 2018.

Challenges Encountered for UAV-Based Surveys

The use of the UAV overpopulated areas presented several obstacles. Overcoming these obstacles will test the viability of the vehicle's use in urbanized areas. For instance, flying directly over people who are not protected by a structure or stationary vehicle is prohibited. Accordingly, best practices dictate that the UAV is piloted to avoid flying over homes and yards where the public and or residents could step directly under the UAV. This risk can be mitigated by pausing the flight or assuming manual control of the UAV. Therefore, generally, the vehicle should be piloted to avoid flying directly over uncontrolled areas, such as private property, without prior coordination with landowners and residents. As safety measures to flight systems and hardware (such as parachutes) advance, restrictions may be relaxed or waived in specific instances.

Flying in coastal California will also be challenging because of the presence of airports, which require special clearance when flying near them. This means extra planning and coordination will be required when flying assessments within 5 miles of major airports. With the incorporation of Low Altitude Authorization and Notification Capability (LAANC), UAS activities in controlled airspace at or below 400ft are much more logistically feasible. With LAANC one can gain approval to fly within subsets of controlled airspace, compatible with UAS based trash assessments, within seconds of submitting your proposed flight activities from a mobile device. In other cases, approval is necessary from the Air Traffic Control. As technology and regulation continue to advance to accommodate reasonable UAS use within the United States, UAS based monitoring may become more widely feasible.

Tree canopy cover may also occlude the UAS based methods by obfuscating the ground and trash present. This often depends on the season relative to deciduous trees.

All of these factors limited and constrained the use of this technology, but where appropriate, the team conducted surveys and reported on the usefulness of the exercise. Figure 14 shows Larkspur Creek, where a combination of trees and private property constrain the survey area. In some cases, sites can be sampled during times of the year when canopy is minimized. Additionally, permission can be sought for flying in areas limited by private property. Even given these constraints, imagery remains a useful tool.

Since site conditions change and flight regulations adjust at a fast pace, some of these rule-based obstacles may change in the near term, perhaps even over the course of this project. The project team monitored changes and made adjustments as necessary.

Figure 29. Larkspur Creek, orthomosaic image, January 18, 2018.

Post-Processing

3DR SiteScan software was used to capture the individual images uploaded by the pilot, orthorectify the images, and assemble them into an orthomosaic. Once prepared in this way, a number of other ancillary products were produced that may inform subsequent analysis, including:

- Contours
- Elevation models (Digital Elevation models and Digital Terrain Models)
- Hillshades
- 3D point clouds
- 3D models

Figure 30. Example UAS photo featuring trash in Wildcat Creek.

Analysis

Following the post-processing stage, analysis was chiefly two-fold: manual and automated.

Manual Analysis

The team enlisted the aid of three trash assessment practitioners who were trained in the BASMAA and SMC methods. They performed “virtual assessments” of the site using only the available imagery, adhering to all applicable guidance as described above. The practitioners were permitted to scour the images.

The team recorded the results and measured the variability of their assessments (precision), as well as how closely they compare to the on-site assessments (accuracy).

Automated Analysis

The team also applied machine learning algorithms to test the viability and practicality of applying these new tools. The team hypothesized that, under certain circumstances, machine learning may be used to accelerate the assessments, thereby potentially expanding the geography and time period surveyed.

The machine learning algorithm is based on TensorFlow, a commonly used computational engine for these tasks. It is a form of a convolutional neural network (CNN) that leverages large datasets to determine patterns.

Our site surveys formed the basis of analysis for the CNN that was charged with the following:

- Identifying trash in the image by individual objects (presence/absence)
- Quantifying the volume of trash in the image, overall

- Depending on the outcome of early optical tests: quantifying the volume of plastic in the image, as distinguished from other forms of trash

By comparing the results of manual analysis to automated analysis, we determined whether there is appreciable difference in performing one vs. the other. Furthermore, with the benefit of on-site extractions, we will have an absolute measure of trash volumes as a basis of additional comparison.

Trash Count Comparison at Varying Flight Elevations

In order to assess the feasibility of detecting trash in a manual and/or automated manner we tested our ability to detect and classify types of trash at different flight elevations.

For this experiment we set out a 2x2 grid of 10’ squares marked with PVC pipe. Each square was numbered in a way that would be visible in UAS based imagery collected. Each square was inspected to ensure no debris was previously within each of the squares and then seeded with trash items. Included were tobacco trash items such as vape pens, cartridges, vape juice bottles etc. Each square was photographed on the ground for validation purposes. We then conducted flights over the quadrant of squares at 100ft, 60ft, and 30ft elevations. A ground based tally/count to identify the type and quantity of trash in each square was then carefully conducted. The volume of trash in each quadrant was then estimated using a five gallon bucket.

Once the field work was completed, the counts/tallies of trash for each quadrant was recorded. Then one of the field crew carefully annotated trash in the 30ft elevation image using LableImg software while referring to tally counts and classifications, distinguishing “tobacco trash” from “non-tobacco trash”. This dataset was used as the gold standard for comparison to what trash was present in the imagery.

A third party, that had not participated in any of the field work or exposed to recorded tallies, then used LableImg software to annotate all observed trash visible in the 100ft elevation imagery as “tobacco trash” and “non-tobacco trash”. The same individual then repeated this step for the 60ft elevation imagery, without referring to the 100ft imagery, and then again for the 30ft elevation imagery, without referring to the imagery/annotations taken at 100ft and 60ft elevations.

True counts that were taken at ground level were then compared to counts made from the 100ft, 60ft, and 30ft images for each quadrant.

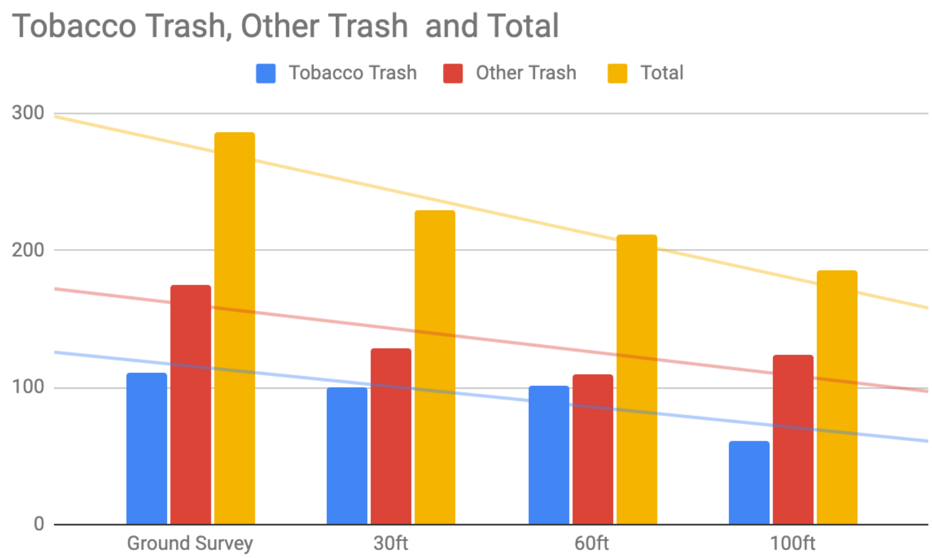


Figure 31. Counts of trash (totals and by type) at 30ft, 60ft, and 100ft elevations compared to Ground Survey Counts.

This analysis shows that although counts are not as accurate from the air, a significant amount of trash on the landscape can be detected from UAS imagery. Furthermore this analysis helped to identify that 100ft to 60ft elevation flights, with a comparable sensor, are suitable for trash detection. Although imagery taken at 30ft provides potentially more accurate counts of trash, 30ft elevation flights and lower are not always feasible in the natural environment due to trees, telephone poles and other tall structures. Trash counts conducted using imagery taken at 60ft and 100ft elevations were similar, however it's interesting to note that classification of type becomes less. Based on this assessment, subsequent flights were conducted between 60ft and 100ft elevation (with a preference to closer to 60ft) when possible and safe.

Machine Learning

Goals

The machine learning work described in this section is meant to augment UAS methods for trash monitoring, particularly to increase the temporal and spatial scope of trash surveys while minimizing time and labor costs. Ideally, such a method could produce a volume estimate and or tallies for the amount of trash at a given site. Considering the novelty of applying this type of technology to aerial imagery, this work should be evaluated with its exploratory nature in mind.

Description of methodology

The sections below describe fundamental aspects regarding how we chose to approach this object detection task. We start with a brief overview of object detection and machine learning pipelines, describe the imagery used in this project, how we prepared and readied data, then explain why we chose convolutional neural networks as our primary algorithm, finally we describe transfer learning and how it applies to our model.

Novel Method Background

Object detection algorithms have been a significant area of study since the 1990s. At the time, object detection depended largely on feature selection methods meant to enhance information in order to categorize items within visual data according to a pre-defined semantic schema. For example, targeting visual cues within scale space to create local scale-invariant features[3]. While similar methods are employed to this day, they require much work to identify and characterize features.

In tandem, neural network (NNs) research has blossomed over the last fifteen to twenty years. They offer the unique ability to function similar to human brains and as a consequence can identify complex patterns within a wide range of data types (e.g. visual, audio, tabulated, etc…). Neural networks can be composed of a wide variety of architectures, each tuned to specific types of data. As of this writing, deep convolutional neural networks (DCNNs) have shown to be very effective for tackling problems requiring visual data sets, specifically object detection within images.[4] It is for this reason we chose to use a DCNN architecture for our application fairly early in the planning process.

DCNNs constitute a vast field of research within the machine learning space. A true primer would be outside the scope of this document; however, there are some key concepts important to outline in order to understand our methodology and analyses. DCNNs consist of huge networks of interconnected layers, each of which serve a fundamental role in analyzing a given data point (or image in our case). Each layer consists of multiple nodes which are the fundamental building block of neural network layers.

We have chosen to use a "feed forward" style DCNN, which means that the training process for our network consists of a set of weighted layers, tuning parameters (i.e. hyperparameters) and visual input. As new visual input is presented to the network, it runs through a series of predictions. Depending on the performance during the prediction phase, it will then update weights for each layer using values we've set for hyperparameters (i.e. optimizers).[5,6]

Dependency on imagery

It's important to note that DCNNs can be applied to a wide variety of data types; however, for our application we will focus solely on RGB color image data obtained from UAV flights.

Preparations

Perhaps the most time-consuming and pivotal part of our work was creating an annotated data set. At the onset of this research, we could not find labelled, aerial imagery depicting trash. Without an initial dataset, the scope of our work had to expand to include the creation of one. This meant that we had to

---

[3] https://www.cs.ubc.ca/~lowe/papers/iccv99.pdf

[4] https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[5] https://arxiv.org/abs/1901.06032

[6] https://arxiv.org/ftp/arxiv/papers/1901/1901.06032.pdf

establish a workflow with which we could draw bounding boxes around example trash imagery and ensure we could format that information for presenting to the TensorFlow framework.

To accomplish this, we identified three trash classifications (unknown, plastic, and not plastic) and annotated over 30,000 images taken over the course of six surveys. Our choice to use three classifications was driven by two factors; more classification categories would result in increased effort to categorize items and advice offered by Ben Woodword, a computer vision expert with CVision.[7]

We chose categories prior to annotating data. As a result, we were not sure how many samples had been collected for each category and have an unbalanced data set. We chose to center our categorical schema around plastics because, at the time of this writing, plastic trash accounts for a large volume of trash found in the environment and poses a significant threat to our waterways, particularly when broken down into microplastics. Our assumption was that we might have need to focus on plastic detection before other types of trash.

For the annotation process, we chose LabelImg,[8] a free open-source annotation software which supports multiple annotation standards. Using LabelImg, we annotated roughly 30,000 samples using the PASCAL VOC2012 standard.[9]

Our annotation efforts yielded roughly 30,000 annotations across all six surveys. Table 4 shows the number of samples we obtained for each category and Figure 32 contains a bar chart created from these totals. Our samples are heavily biased towards trash unknown, which can heavily bias the dataset and result in a disproportionate number of detections labelled as "trash unknown." Figure 33 shows an example of the LabelImg interface during the annotation process. The image shows survey imagery with small boxes, connected via green dots, which serve as annotation boundaries. The right hand column shows a list of labels and list of images for the full survey. The left hand column shows tools that can be used to create and edit annotations.

Table 5. Trash counts within annotated dataset.

Category	Counts
Unknown	28,404
Plastic	1,493
Not Plastic	738
Total	30,635

[7] https://cvisionai.com/

[8] https://github.com/tzutalin/labelImg

[9] http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

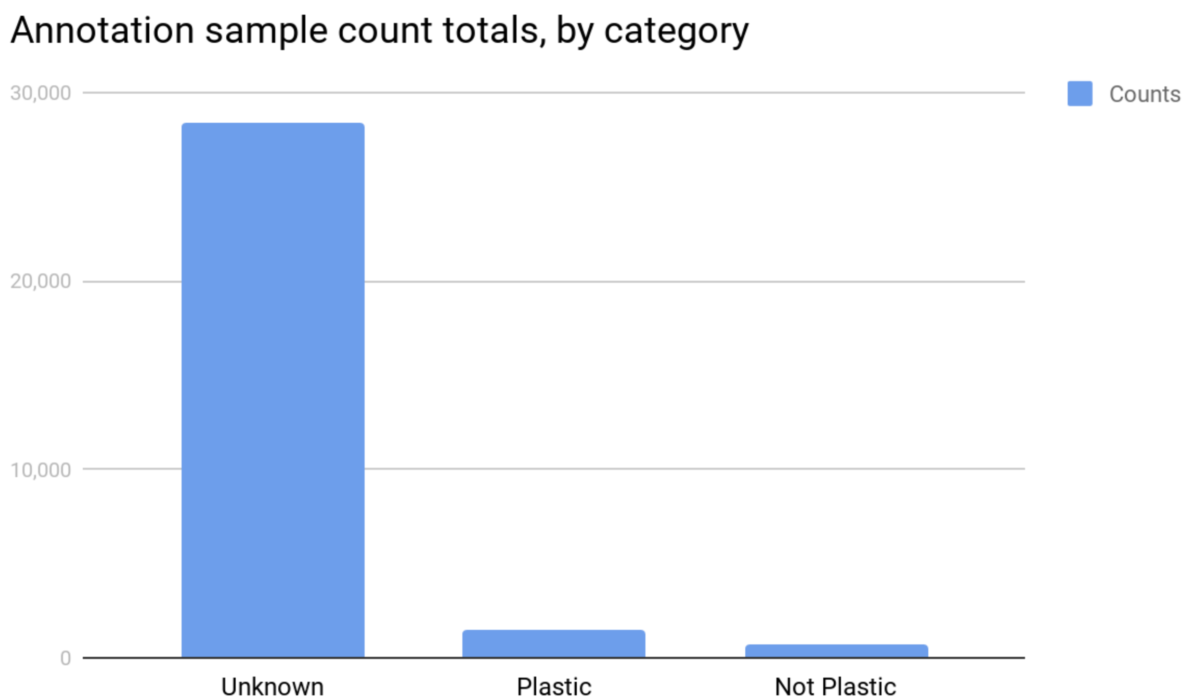


Figure 32. Annotation sample counts totals, by category

Figure 33. Annotations made using LabelImg

Justifications for Algorithm and Framework Selection

Given time and resource constraints we decided early on to leverage deep convolutional neural networks (DCNNs) for detecting trash within wetland environments. DCNNs have the ability to identify patterns within datasets. Other machine learning models, such as Support Vector Machines (SVMs) or Random Forests (RFs) depend on a combination of pre-processing, tabularization, and post-processing steps. While we could theoretically leverage unsupervised learning methods such as clustering or maximum likelihood estimation to automate pattern recognition as part of a process using these other machine learning methods, exploring such methodologies would be costly. We had to ensure we had enough time to handle workflow issues such as creating the dataset, running training computations, verifying accuracy, etc.

It was also important to use a mature software platform and avoid the initial cost of training a neural network from scratch. TensorFlow as a natural choice and includes an object detection API[10] along with a pre-trained model from TensorFlow's detection model zoo[11] as a starting point for our model training.

The TensorFlow object detection API contains workflows for streamlining common object detection workflows. Tools include but are not limited to reading PASCAL VOC2012-formatted bounding box annotations, inputting associated visual data into a neural network model, perform prediction iterations, update neural network layer weights, and other tasks required to build inference models meant for object detection.

Transfer Learning Process

We chose to train our model using the Faster RCNN[12] Inception Resnet v2[13] Atrous[14] COCO[15] architecture, available on the TensorFlow detection model zoo github page, as the starting point for our model training. This architecture runs quickly and can be used to detect relatively small objects.

Since trained models available in the detection model zoo are meant to serve as a foundation for further training, they are pre-trained just enough to integrate into a training workflow which focuses on training the networks final, fine-tuned layers which are meant to differentiate slight variations of visual patterns to differentiate objects. These models include default pipeline configurations meant to serve as conservative hyperparameter values which can achieve decent results given enough training time.

---

[10] https://github.com/tensorflow/models/tree/master/research/object_detection

[11] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md

[12] https://arxiv.org/abs/1506.01497

[13] https://ai.googleblog.com/2016/08/improving-inception-and-image.html

[14] https://arxiv.org/abs/1706.05587

[15] https://cocodataset.org/

As of March, 2020 we obtained access to resources which have let us experiment with model sensitivity to various training parameters; however, we simply have not had enough time to identify a set of parameters which performs better than the model we initially started training last year.

Testing Process

There are a few ways to evaluate the models accuracy, we've chosen the following three methodologies to examine model efficacy:

1. Accuracy and recall numbers reported by the training process.
2. Trash count comparison leveraging a controlled trash survey.
3. Trash count comparison to survey tallies as reported via the SMC methodology.
4. Volume assessment comparison between volumes reported from surveys with those obtained the object detection algorithm.

Model Accuracy Reported During the Training Process

As described in the Novel Method Background section, during the training process the network performs a series of predictions. These predictions are compared to actual data and in turn drive how the network is updated upon the next iteration.

The Faster RCNN arduous COCO model model we chose, leverages many metrics for evaluating performance categorized by object size (i.e. large, medium, small). Mean average precision (mAP) and average recall (AR) are reported for each size as well as a value aggregating across all sizes.

Mean average precision (mAP) measures how well the model fits annotated objects during the training process. For example, mAP=1 would mean that during training iterations the model is able to fit an exact square matching bounding box dimensions for corresponding annotations.

Average recall (AR) measures how well the model identifies existing annotations during the training process. For example, AR=1 would mean that the model is able to detect every annotation with a bounding box that overlaps roughly 50% or more with the annotation bounding box.

While we've conducted a series of experiments using various sets of hyperparameters, our best performing model is also the model that has been training off and on over the last year. That being said, our model has not performed very well by these measures. Figures 34 and 35 show example graphs for mAP and AR as of the time of this writing. You can see the highest mAP value is around 0.045 and the highest AR. These are very low values and would indicate the need for further testing.

Figure 34. Mean Average Precision (mAP) v.s. Training steps

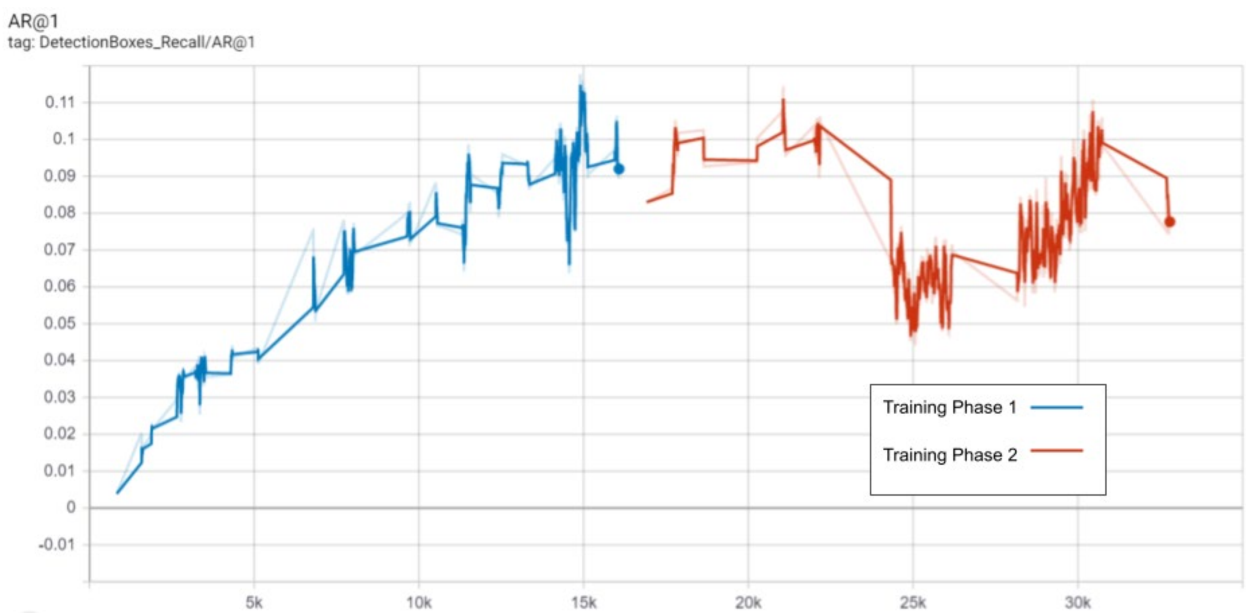


Figure 35. Average Recall vs. Training Steps

Trash Count Comparison to Controlled Trash Survey

While metrics reported during the training process have been fairly low, we've decided to test our best performing model against a controlled trash survey. The goal here would be to calculate an accuracy metric more reflective of real world applications.

For this survey we created a 2x2 grid and placed items of trash into each quadrant. We conducted drone flights at 30ft and 60ft altitudes. SFEI staff visually inspected this imagery and collected totals for each quadrant. We then used our best performing model to detect trash in each quadrant as well.

Human visual detections give us a sense of what kind of accuracy is achievable using drone imagery. It then serves as a basis of comparison for evaluating automated detections using our trained model.

Table 5 aggregates totals reported by two SFEI staff and the machine learning model for detection confidence greater than 0.5 (50%) and 0. Figure 36 shows this data in bar chart format as well. Perhaps most eye-opening is that actual counts, conducted on the ground, differ quite a bit when compared to both human tallies. Similarly, human based tallies are fairly different, which implies that we need to incorporate further human testing to get a true base-line for comparison between model predictions and human-based aerial surveys.

It is obvious that with no detection threshold (0) and 50% detection threshold (0.5) the model is an order of magnitude less accurate than actual on-the ground tallies. On the one hand, this implies the model, at its current performance, is fairly conservative at identifying a given object as trash. Figures 37 through 40 show inference output from our model with confidence level greater than 0.5.

Table 6. Trash counts for controlled survey

Quadrant	Human 1	Human 2	Model (> 0.5)	Model (> 0)	Actual
1	4	7	3	5	12
2	7	12	2	6	20
3	6	14	3	8	28
4	3	19	1	2	22

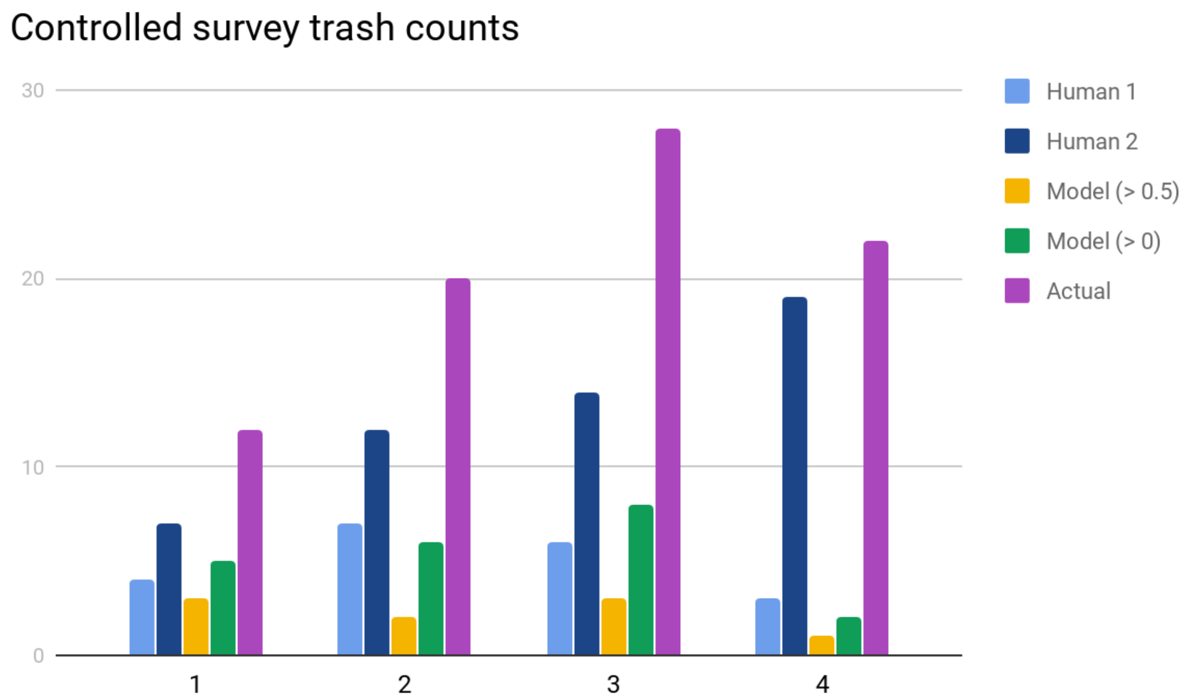


Figure 36. Detection comparison results

Figure 37. Quadrant 1 inference output (confidence > 0.5)

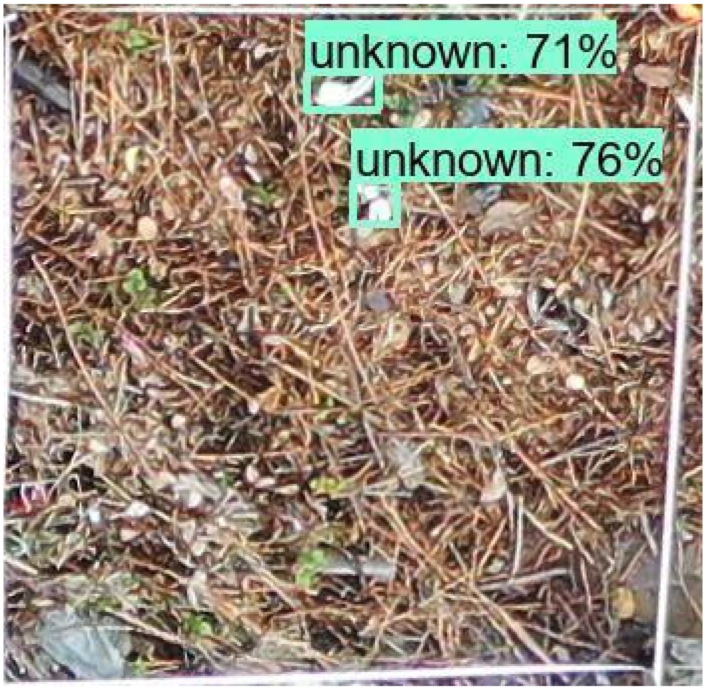


Figure 38. Quadrant 2 inference output (confidence > 0.5)

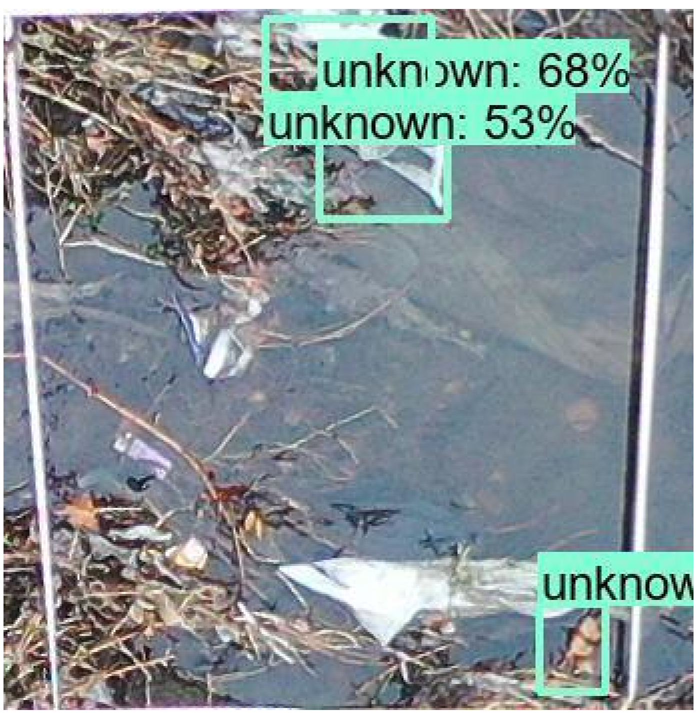


Figure 39. Quadrant 3 inference output (confidence > 0.5)

Figure 40. Quadrant 4 inference output (confidence > 0.5)

Trash Count Comparison to Survey Tallies

We also wanted to evaluate how well the model performs against survey results obtained from ground based survey efforts outlined in this report. Acknowledging that human remote counts have captured significantly less trash when compared to on the ground surveys, it's realistic to assume aerial based detection methods will most-likely undercount trash totals when compared to ground-based surveys.

While conducting this comparison we encountered the following difficulties:

- The model frequently detects rocks as false positives.
- It seems even within the Bay Area wetland terrains differ enough that some environments might produce more false positives than others given our initial training set.
- The SMC Tally method reports totals from ground based surveys which makes it difficult to rectify trash detection locations from aerial surveys with ground data.
- Conducting detection over orthorectified imagery seems feasible, which would avoid the need to create an algorithm for identifying duplicate detections across individual survey images.
- While there are a significant number of false positives, detection totals do seem to correlate with tally totals for two surveys based in Contra Costa. Detections for the Livermore-based survey do not correlate with count totals.

Below are survey descriptions accompanied by model detection totals with qualitative false positive assessments and some detection output examples.

Contra Costa - Oak 01 - 100ft

Survey contained little to no trash, some objects easily viewed from aerial imagery. Upon visual inspection, roughly half of the detection > 0.5 are false positives.

* Date: 8/14/2018
* SMC Tally Total: 31
* Model detections > 0.5: 14
* Model detections > 0: 24

Example false positives:

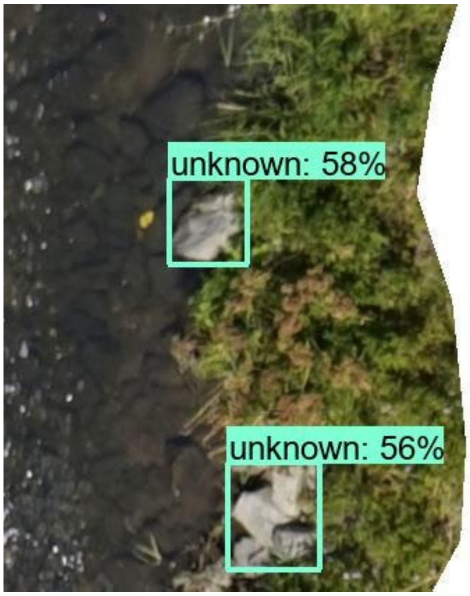


Figure 41. Example False Positives from Contra Costa - Oak 01 - 100ft

Example true positives:

Figure 42. Example True Positives from Contra Costa - Oak 01 - 100ft

Contra Costa - CLA_1 - 100ft

Survey contains a fair amount of trash, many objects visible from aerial imagery. Upon visual inspection, roughly half of the detection > 0.5 are false positives.

- Date: 8/14/2018
- SMC Tally Total: 276
- Model detections > 0.5: 54
- Model detections > 0: 105

Example False Positives:

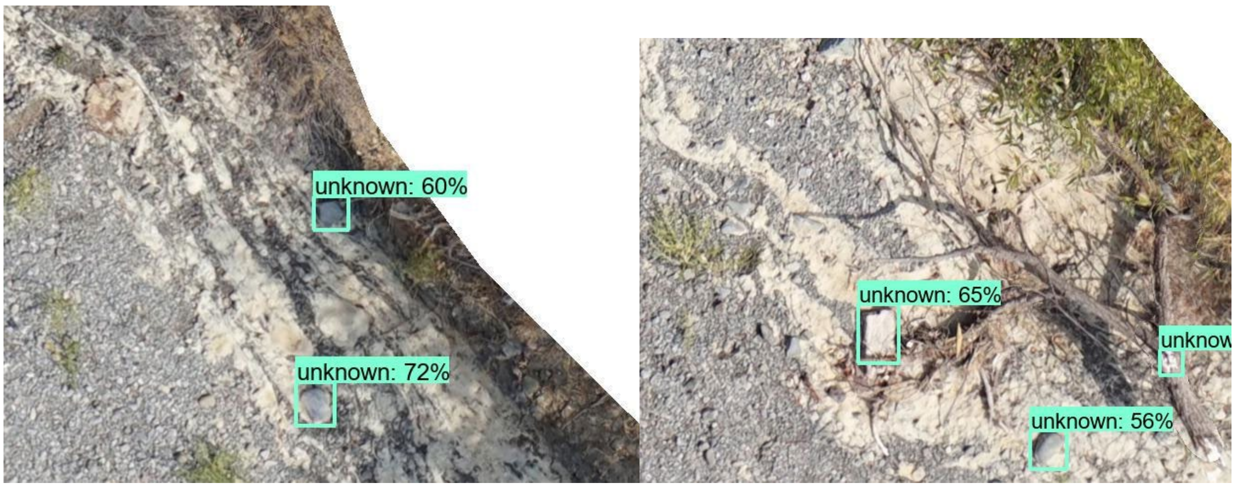


Figure 43. Example False Positives from Contra Costa - CLA_1 - 100ft

Example True Positives:

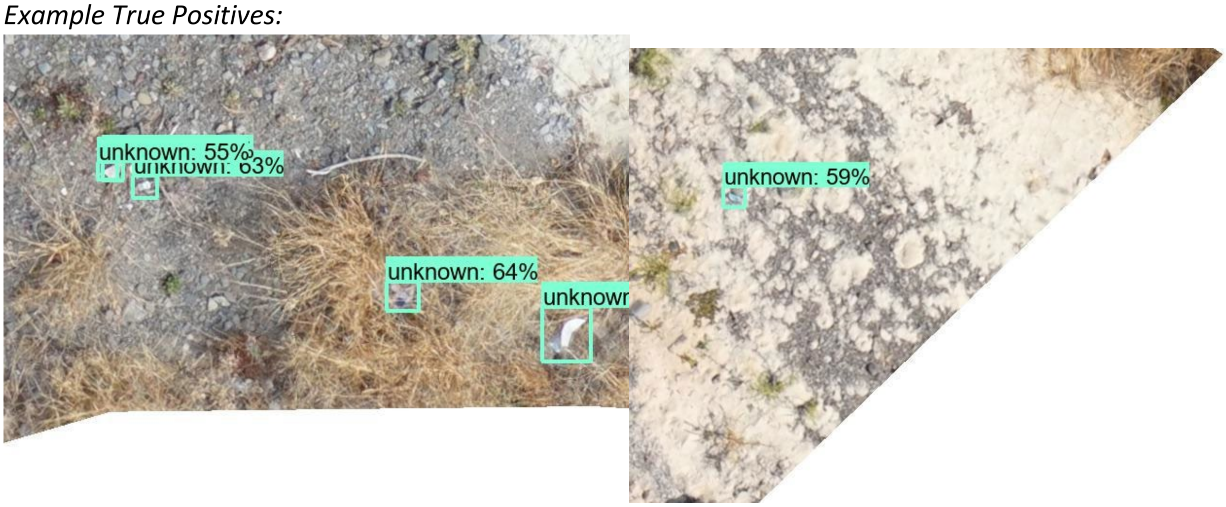


Figure 44. Example True Positives from Contra Costa - CLA_1 - 100ft

Livermore - LIV_ALP_1 - 100ft

Survey contains a medium level of trash, objects are fairly difficult to identify visually from aerial imagery. Majority of automated detections are false positives.

- Date: 7/25/2018
- SMC Tally Total: 62
- Model detections > 0.5: 129
- Model detections > 0: 258

Example False Positives:

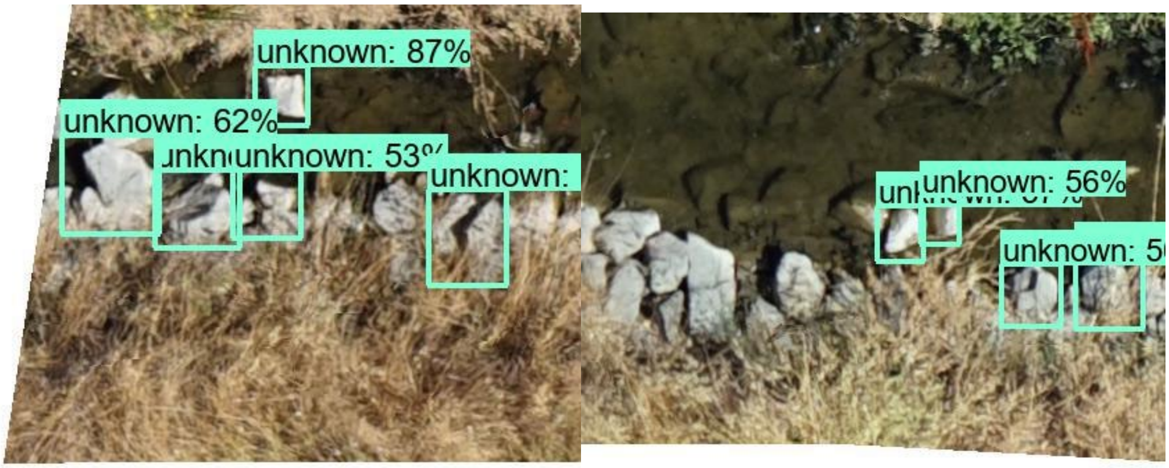


Figure 45. Example False Positives from Livermore - LIV_ALP_1 - 100ft

Example True Positives (with some false positives):

Figure 46. Example True Positives from Livermore - LIV_ALP_1 - 100ft

Acacia Ditch - Fairfield_AcaciaDitch - 100ft

Survey showed the site contained lots of trash; however, not much visible from remote imagery. Reviewing predictions shows few potential false positives and many potential true positives.

- Date: 12/19/2019
- ETAP Survey Tally Total: 849
- Model detections > 0.5: 40
- Model detections > 0: 61

Example False Positives:

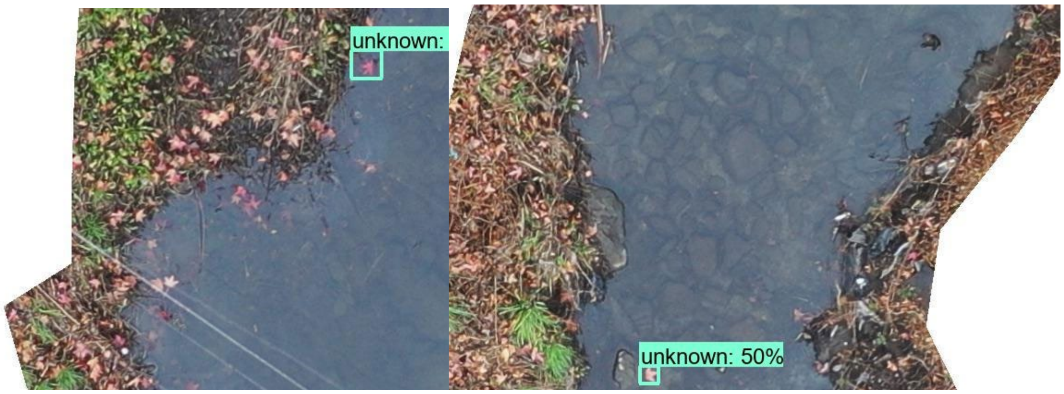
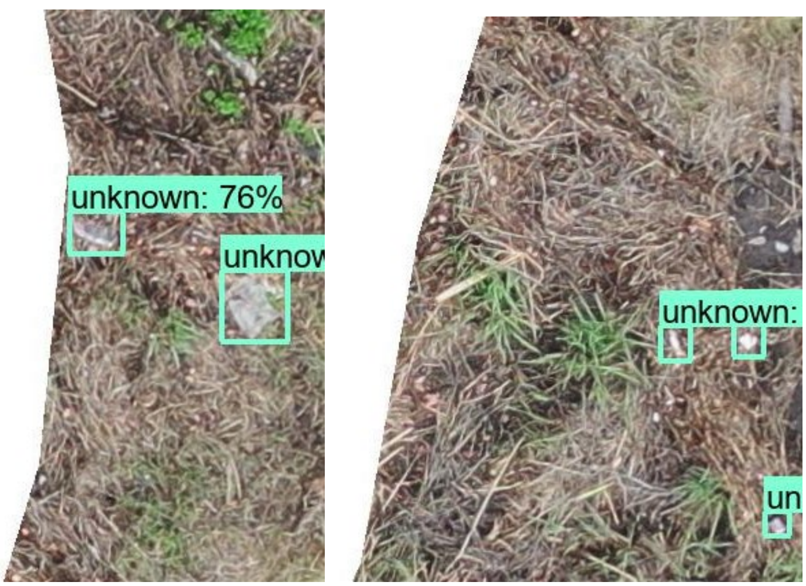


Figure 47. Example False Positives from Acacia Ditch - Fairfield_AcaciaDitch - 100ft

Example True Positives:



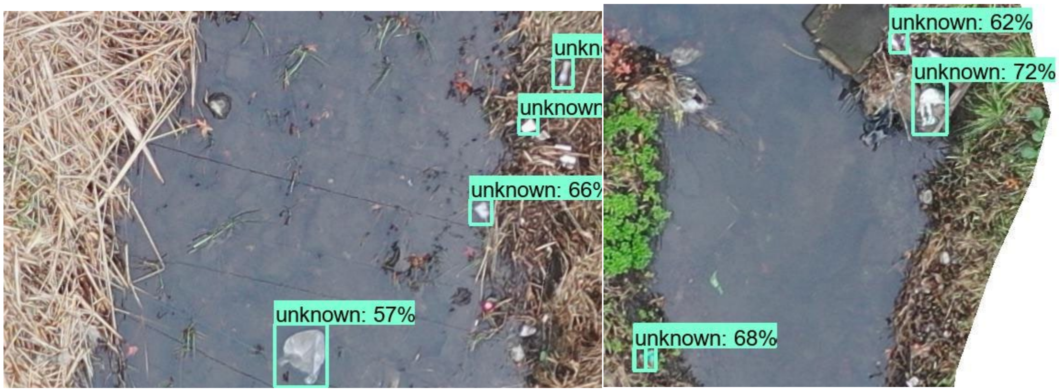


Figure 48. Example True Positives from Acacia Ditch - Fairfield_AcaciaDitch - 100ft

Verde Elementary Upstream - VerdeElementaryUpStrm - 85ft

Survey flight was slightly lower than the typical 100ft. Survey tally total indicates a moderately trash dense site. Reviewing predictions shows few potential false positives and many potential true positives.

- Date: 10/16/2019
- ETAP Survey Tally Total: 207
- Model detections > 0.5: 22
- Model detections > 0: 35

Example False Positive:

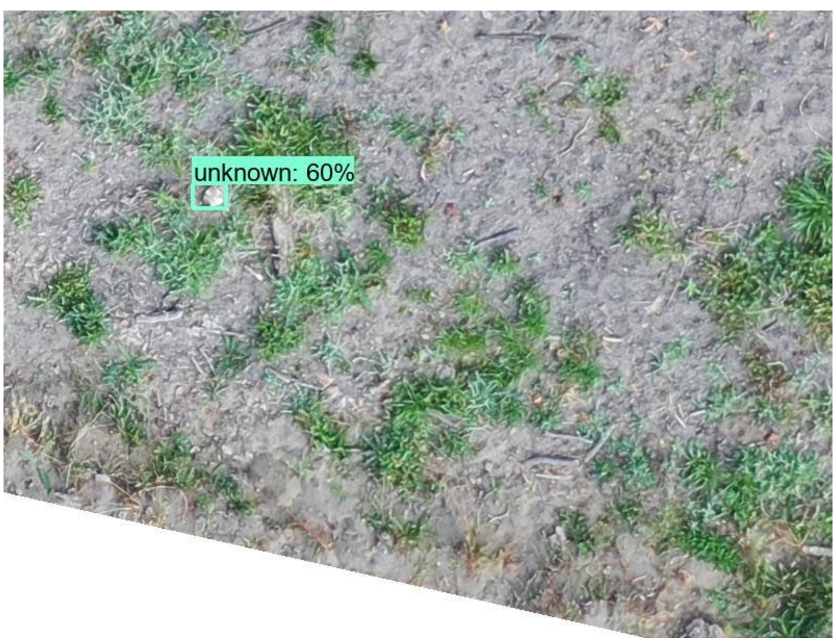


Figure 49. Example false positive from Verde Elementary Upstream - VerdeElementaryUpStrm - 85ft

Example True Positives:

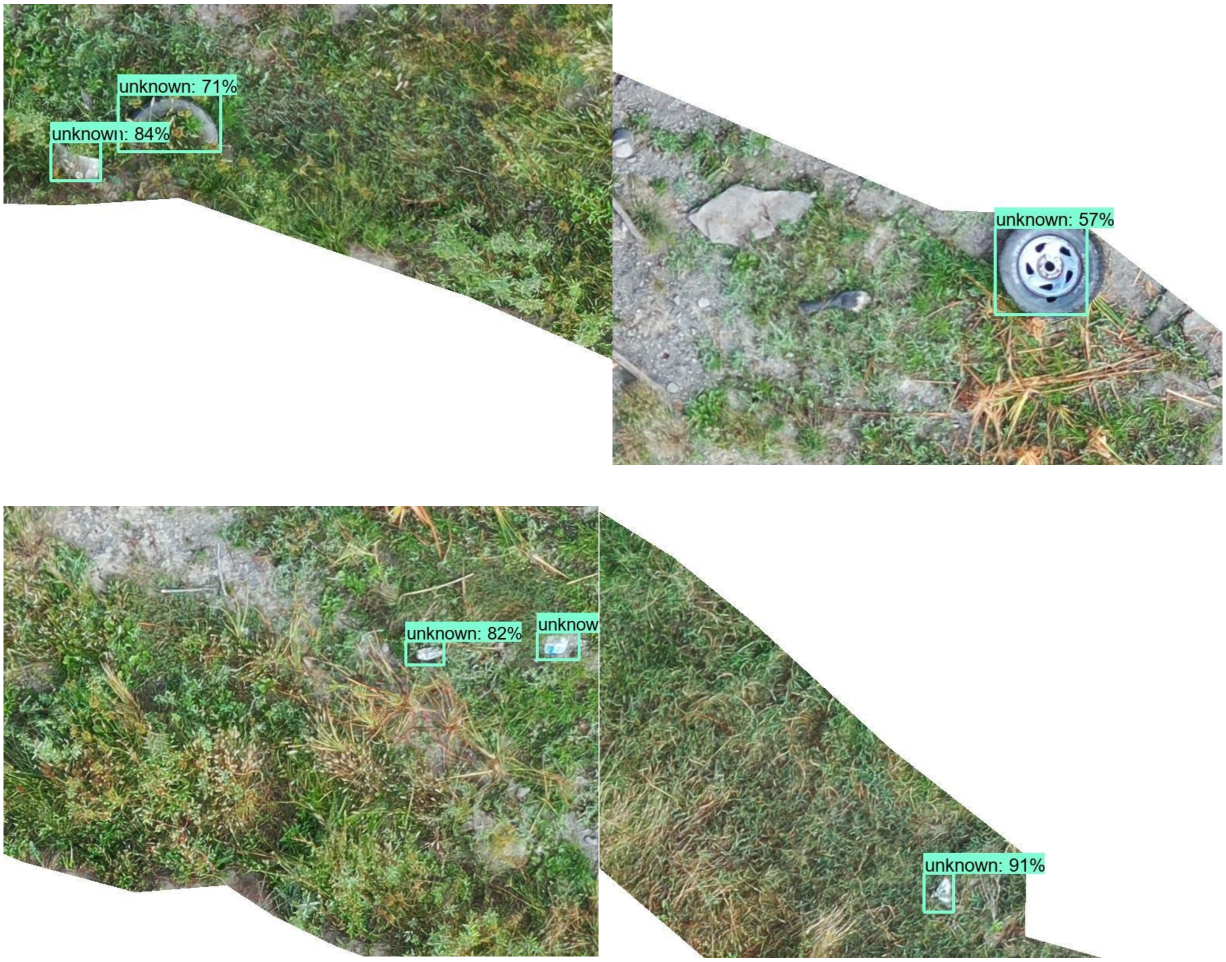


Figure 50. Example true positives from Verde Elementary Upstream - VerdeElementaryUpStrm - 85ft

Ohio Ave Lower - OhioAveDownstream - 70ft

Very similar to Verde Elementary Upstream, this survey was flown below our usual 100ft altitude. Was lightly trash dense based on the ground-based survey and also had few visible false positive predictions. There were also some potential non-plastic and plastic predictions identified, which shows the model could potentially leverage these categories despite the training data imbalance.

- Date: 06/12/2020
- ETAP Survey Tally Total: 256
- Model detections > 0.5: 42
- Model detections > 0: 53

Example False Positive:

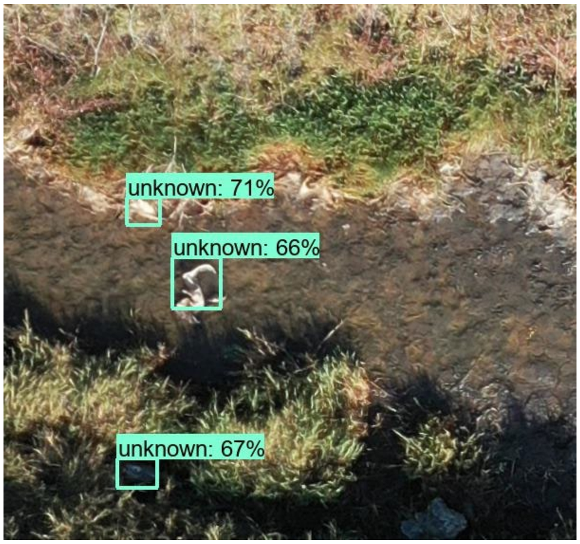
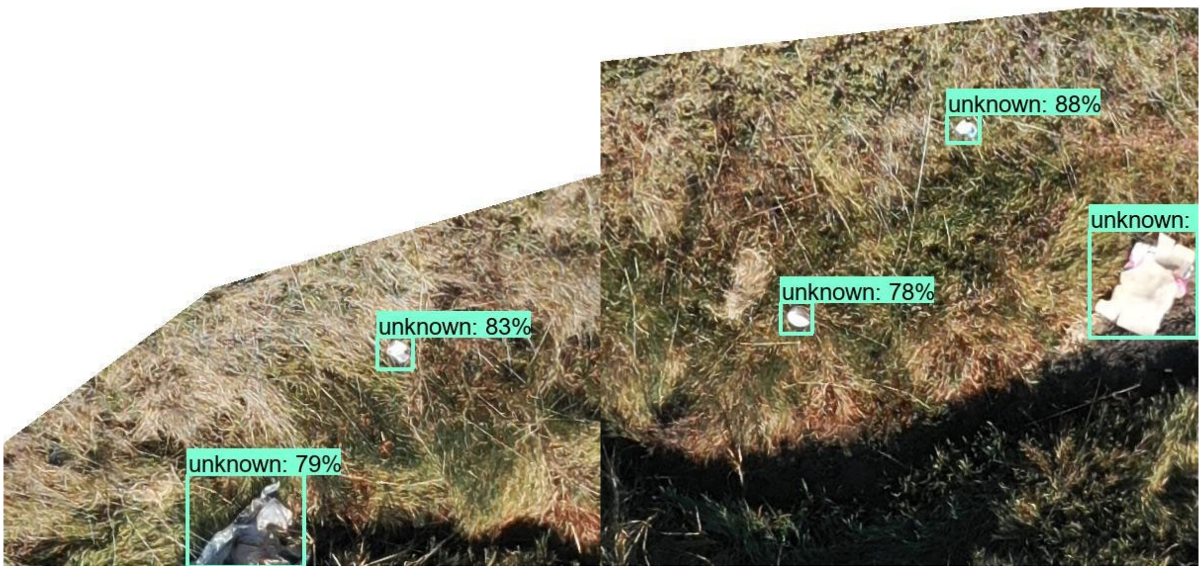


Figure 51. Example false positive from Ohio Ave Lower - OhioAveDownstream - 70ft

Example True Positives:



unknown: 88%
not-plastic: 70%

plastic: 62%
unknown: 72%
unknown: 72%

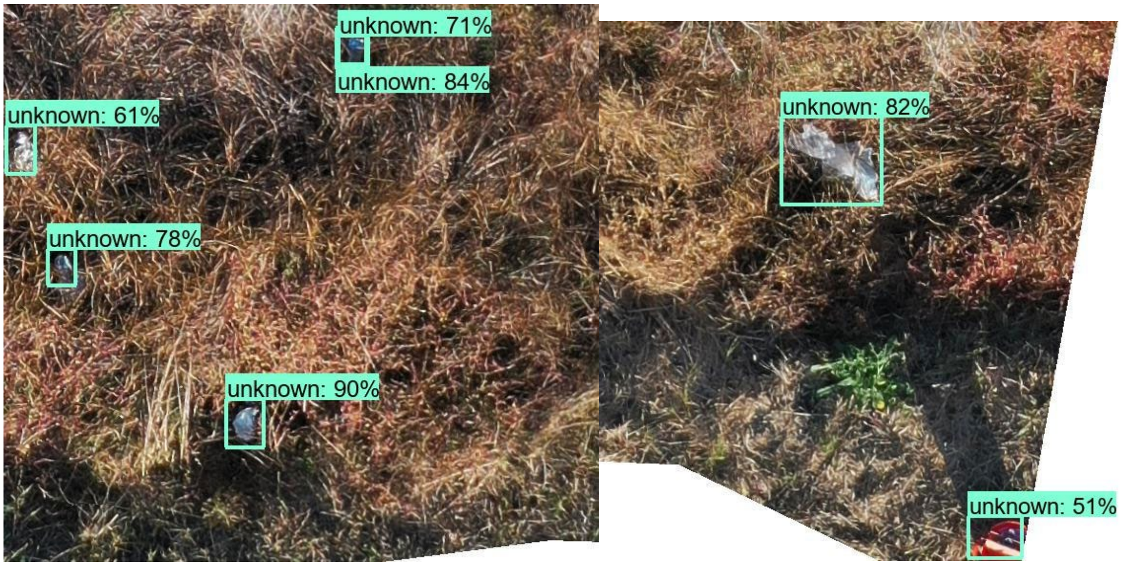


Figure 52. Example true positives from Ohio Ave Lower - OhioAveDownstream - 70ft

Volume Assessment Comparison

Since our primary goal is to develop a methodology for assessing trash volume using a combination of drones and machine learning, we felt it important to see if we could identify a methodology for predicting trash volume. A remote-based volume assessment methodology would require sensors capable of light detection outside the visible spectrum coupled with pre- or post- processing steps. Such a system, while effective, would be outside the scope of our mission to ensure this methodology is affordable. With that in mind, we decided to see if we could leverage a regression analysis to calculate a formula for predicting volumes based on total trash tallies. The expectation would be that volume estimates calculated using totals from model inferences would differ from those calculated using on-the-ground survey totals; however, if those estimates can be correlated, perhaps we could refine a methodology for obtaining more accurate volume estimates.

To accomplish this, we started by using a 2nd degree polynomial regression fit taking trash counts as input (x) and trash volumes as outputs (y). Figure 53 shows a plot and fit for a data set derived from Verde Elementary Upstream, Ohio Ave Lower, and Acacia Ditch surveys.

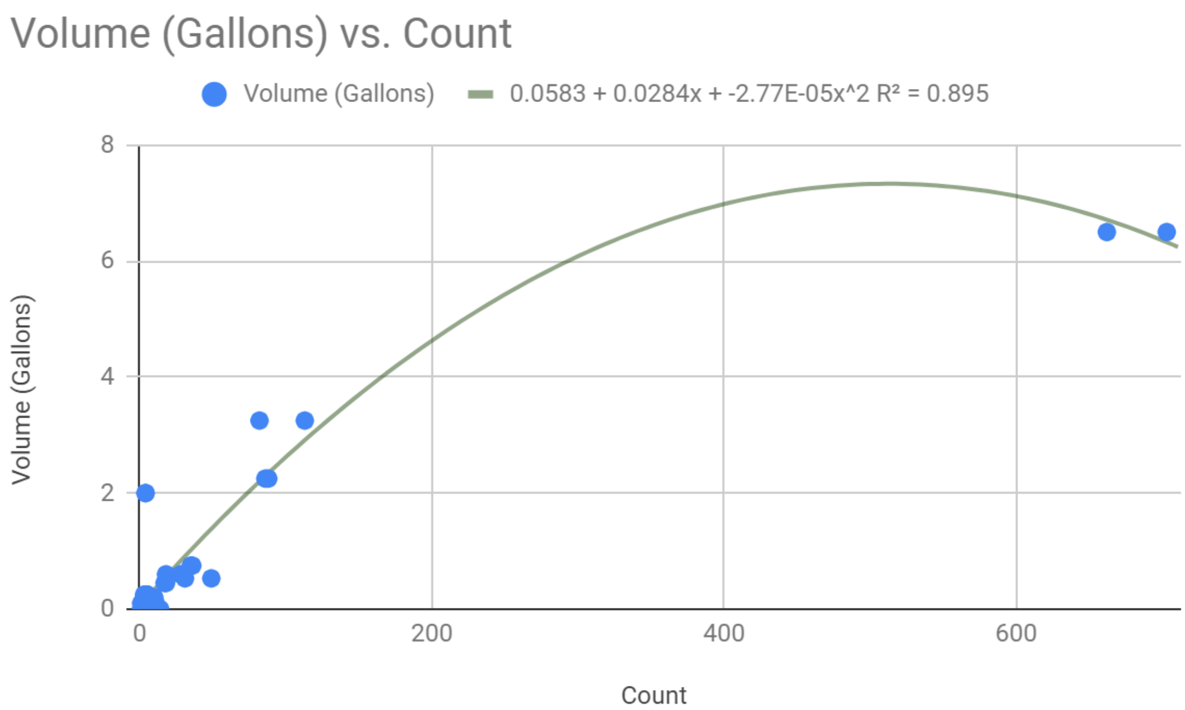


Figure 53. 2nd degree polynomial fit to count vs volume data.

Next, we sought to calculate volume estimates using the fitting formula and compare to survey volumes. As expected, these predicted volumes were consistently less than survey volumes. Figure 54 shows volume estimates for predictions with greater than 50% confidence (>0.5) and any confidence (>0) next to survey volumes. Surveys containing the term (ETAP) represent a reduced area surveyed at that location.

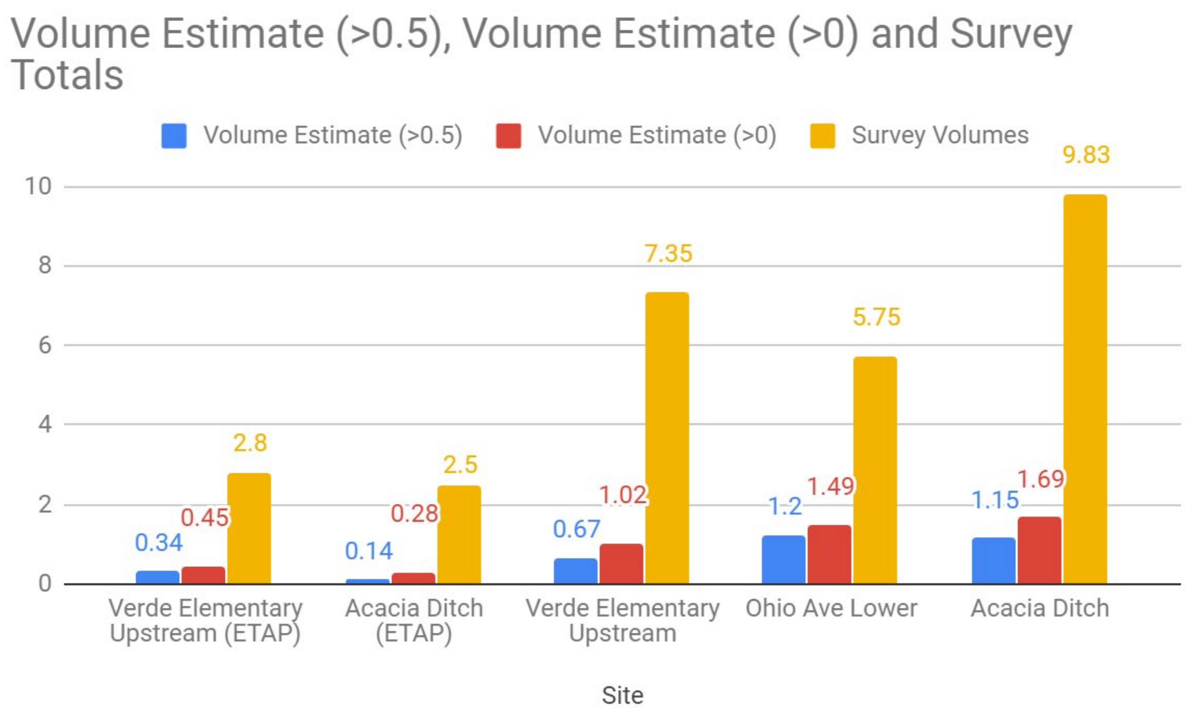


Figure 54. Estimated and survey volumes.

Finally, we calculated correlation values between estimates derived from both confidence thresholds and survey volumes. We found both thresholds yield positive correlation values, 0.79 for confidence > 0.5 and 0.89 for confidence > 0. Both correlation values indicate a potential relationship between these data trends and hint that there may be a way to leverage predicted volume estimates obtained from model inferences as a signal proxy for trash volume.

Lessons Learned

There are many lessons learned from this work that will greatly inform future work. Below is a list of points to highlight:

- The composition of trash within training data, coupled with the surrounding terrain, seem to heavily impact how effective the model will be when conducting trash predictions over novel terrains.
- Since the TensorFlow object detection API can only leverage one GPU at a given time, this becomes problematic when attempting to train models quickly.
- Orthorectified imagery could serve as a viable input when conducting inference, thus mitigating the need to develop an algorithm for detecting duplicate detections across individual survey images
- When we started this work we were unsure if using machine learning to detect trash would be feasible. We believe model inferencing output demonstrates that, while difficult, automated trash detection is possible.
- Due to our imbalanced dataset, most detections are categorized as trash unknown.
- Both Ohio and Verde Elementary surveys were flown below 100ft and yielded promising results. There's a chance lower altitude flights might yield better performance.

Future Opportunities

Perhaps the biggest lesson has been that there are many avenues to pursue when attempting to make trash detection models that could reliably detect trash. Every aspect of the workflow we've outlined in this section has potential for refinement and further investigation. Below are descriptions for potential future work which could improve model performance and get us closer to using this methodology for volume prediction.

Training Set Refinements

It would be a great next step to obtain further annotations and create a training set that is more balanced between categories. This would allow the model to differentiate categories better.

We could also target training sets to fit surrounding terrain better. If we could identify an easy way to characterize terrain types, we could better match sample data and build a model targeting a given location. Ideally we could take multiple surveys at the same site over time and build a model only using data collected from there.

We could also leverage the existing model to conduct a first pass at annotating future datasets and reduce the workload required. This is considered bootstrapping the training set and is commonly done within the machine learning field.

Algorithm Testing and Refinement

Ideally, we would conduct further tests using alternative neural network architectures and their respective hyperparameter spaces. Other architectures might operate better with small objects or differentiate trash specific features.

Ensemble based classification has been shown to improve performance in some instances, it might be beneficial to create an ensemble of models using distinct architectures. We would then aggregate prediction scores across all models when performing inferences over novel data.

Furthermore, there are frameworks beyond TensorFlow which have alternative neural network implementations and architectures. As of this writing two big contenders would be Caffe and PyTorch as they are heavily used in natural science applications. Given enough time, we could also try developing a custom architecture.

Testing and Validation

We believe investing in surveys designed to identify the best a human observer could achieve when categorizing trash from drone imagery would be beneficial for any future work we perform. Considering one of the lessons we learned was that we need to derive an accuracy assessment tuned to our objectives, this will be

> The composition of trash within training data, coupled with the surrounding terrain, seem to heavily impact how effective the model will be when conducting trash predictions over novel terrains.

valuable no matter how we approach other future work. One of the easiest ways to start this would be to have an expert annotate one or more new surveys and use them only for accuracy calculations.

Accuracy calculations derived in this way would give us a better idea of what's possible with automated detections. It would give us realistic metrics to compare count totals and volume estimates with. As noted in the controlled survey, human detection counts can be dramatically different, so it would be important to compare annotations between experts to identify potential errors.

Trash Volume Assessments

As described in our volume assessment comparison, we observed a high correlation between volume predictions and survey volume totals, which supports efforts to explore volume prediction methods.

At a basic level, conducting more surveys with granular volume data would add more data points for our regression analysis and would improve confidence in our correlation values. If correlation values can stay consistently high as more data is added, we could eventually use them as a proxy in real world applications.

We may also eventually leverage training set and model architecture improvements previously described to reduce false positive rates within predictions. This would in turn dramatically improve the reliability of tallies used to calculate volumes. Similarly, we could potentially identify a filter phase in which we filter out false positive counts and ensure volumes are calculated from as many true positives as possible.

Conclusions

The methods evaluated in this report all measured different aspects of trash in the environment. Overall, estimates for accuracy, precision and resources vary (Table 7). Most of them were relatively accurate and precise both within and among groups when training was provided (see individual sections on each method). The methods addressed different types of monitoring questions and should be considered separately when determining which method to use.

Table 7. Comparison matrix for tested methods.

METHOD	ACCURACY	PRECISION	RESOURCES
Visual	Low-Med	Low-Med	👤
UAS	Low-Med	Low-Med	👤👤
Volume	Med-High	Low-Med	👤👤👤
Tally	Med-High	Med-High	👤👤👤👤

Accuracy and Precision - Size Matters for All Methods

In most cases, the size of the trash items impacted the accuracy and precision of the amounts of trash measured. The results for the Extinction Curve using the Tally Method indicate that the smaller trash you have at a given site, the more passes are needed to account for all the trash. In other words, if only a single pass is made at a site, the probability of missing smaller items is increased. This likely holds true for the volume method as well, although the impact of missing smaller items there may be less as the smaller items have much smaller volume measurements. The qualitative method that assesses the site condition based on visual inspection also likely is impacted by the size of trash at a given site, as the smaller sized trash is not always seen and considered when assigning a score to a site. This also holds true for the UAS method, as smaller trash in the range of less than 3 cm is often not seen in the imagery, and even when it is seen, it is much more difficult to identify.

Resources – Costs Vary by Resource Type

The amount of resources needed for each method varies. In general, resources for the Qualitative Assessment are the lowest, with the highest being required for the Tally Method and ETAP. For the Tally Method, ETAP, and the Volume Method most of the resources are in the form of labor as high trash sites require more labor for longer amounts of time. Low labor methods include the UAS and Qualitative methods. Material resources are the greatest for the UAS method initially, as the equipment required (e.g. drone, software, server, etc.) is costly, though coming down. Additionally, any costs put in up front diminish over time with the use of the equipment. Labor in this case is minimal as the time it takes to collect the imagery is less and only two people are required. Post processing costs are also minimal as the bulk of the work is done via machines. The Qualitative Method is the least expensive all around in that it requires the least amount of material resources and minimal time with a crew of two people.

Method Relationships - Using One Method's Results to Predict Another's

The question that has been posed by many is "Can one method be used as a predictor for the results of another method?" This is a tough question to ask as the methods themselves seek to answer different monitoring questions. As was seen in the section on Site Information, the hypothesis that the amount of vegetation would predict the amounts of trash at a given site did not pan out. Meaning that a higher amount of vegetation should predict higher amounts of trash due to entanglement and retention of the trash within the vegetative area. However, this study and BASMAA 2017 showed that this is not necessarily the case.

In comparing the methods to one another we found that the relationship between tallies and the volumes was low (R^2 = 0.459; Figure 55). This is not surprising as the amount of trash can vary greatly depending on the size of the trash items. While you may have a large amount of trash by count the volumes could be low given the size of the trash.

Which Method is Best?

By way of this project, we are often asked to identify the "best" method. The Playbook ideally aims to dispel a notion of the "best." Rather, one might instead consider degrees of suitability to address management and related monitoring questions. Through Table 7, we relate the accuracy, precision, cost, and ease of use for each method. See the companion document to this report, the California Trash Monitoring Playbook for more information on how the methods relate to management questions.

The rapid trash assessments developed by BASMAA offer more coverage and speed at the expense of greater specificity of items, materials, or item counts. The quantitative visual method developed by SMC, on the other hand, is more time-consuming and costly, with smaller coverage and a relatively higher degree of accuracy.

Importantly, however, if you have trash-related questions specific to certain materials, the BASMAA method might only have limited suitability. Whereas a tally method would be optimal to address such questions. If you wish to characterize an overall change in trash load for a given municipality or neighborhood, however, resources might dictate that BASMAA's method would be the practical choice.

It is therefore critical to evaluate the suitability of the method in light of the monitoring question, available resources, and your needs for accuracy, precision, and monitoring target.

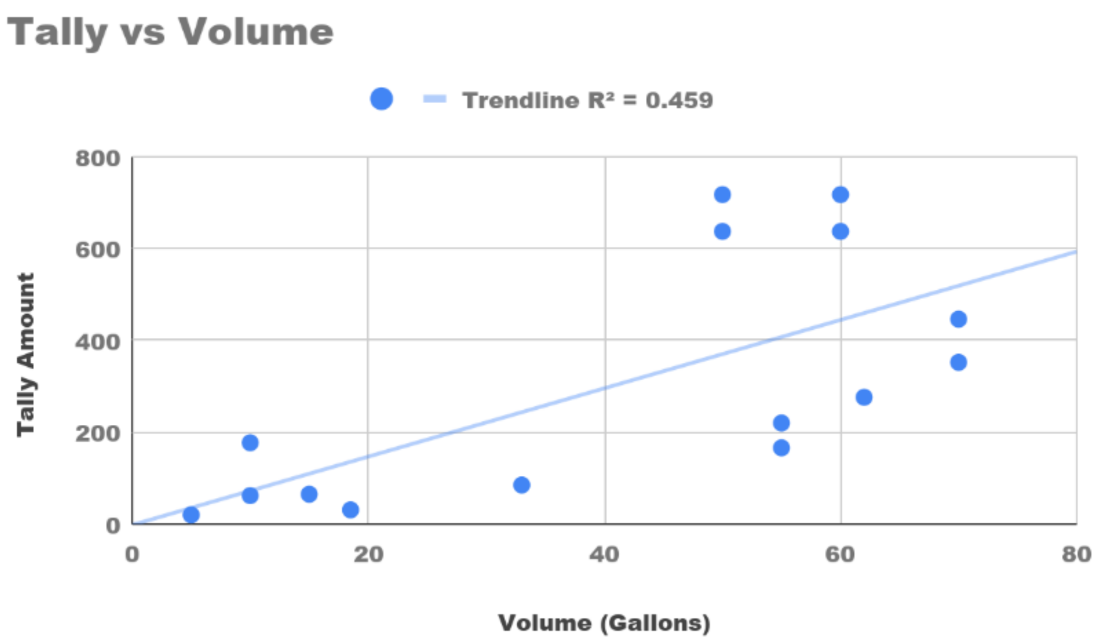


Figure 55. Volume measurements versus Tally Counts by site.

For the total trash count versus the Qualitative Site Condition Score, the relationship was also low (R^2 = 0.473; Figure 56). Again, this is in large part likely due to the variation in the size of the trash at a given site. A site could have large amounts of trash by count but if it is small, the perception of trash in the assessment area would warrant a low site condition score.

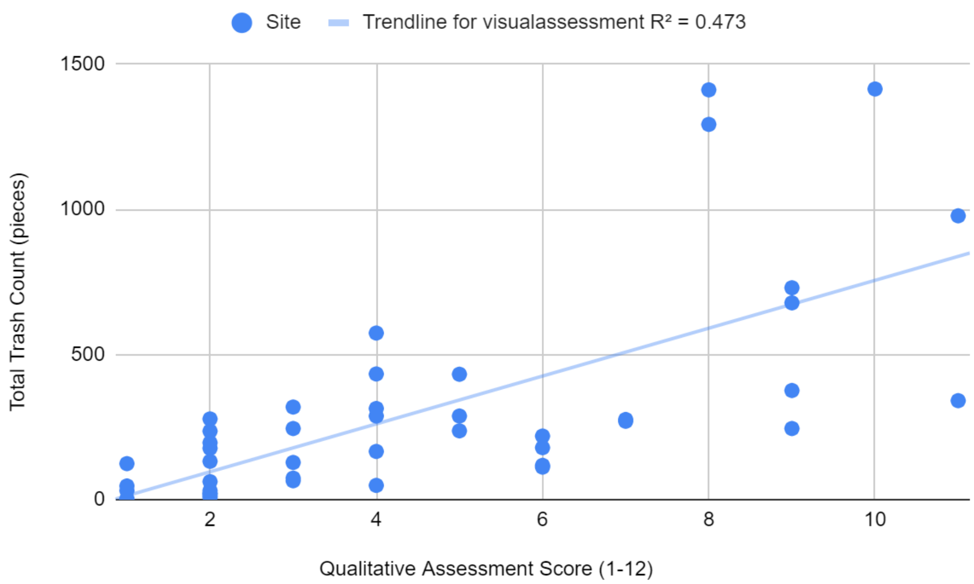


Figure 56. Qualitative Assessment Score versus Total Trash Count by Site.

For the Tally Method versus the UAS method there was much difference in the amounts of trash estimated based on the size as well. The Project Team seeded an assessment site with known amounts and types of trash and measured the amounts of trash using both methods. The differences in counts were found to be greatest when all trash (all sizes) was counted using both methods (Figure 57) and decreased as pieces and subsequently paper were removed. The smallest difference in amounts measured between the two methods was found when both pieces and paper were removed.

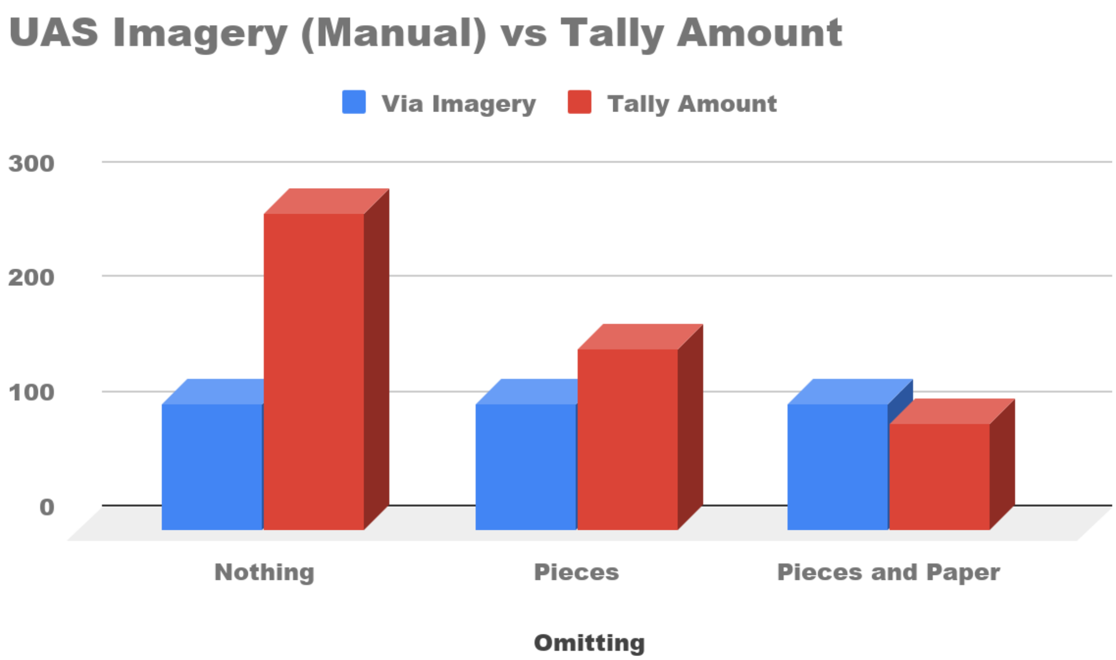


Figure 57. Amounts of trash counted via the UAS and Tally Method. Second set of bars represents counts after removing pieces and the third set represents both pieces and paper.

The one area where the relationship between methods was high was between that of Qualitative Site Condition Assessment Method and the Volume Method (R^2 = 0.894; Figure 58). This is also consistent with what BASSMA found for the San Francisco Bay Pilot Trash Monitoring Project (BASMAA 2020). The thought here is that what the assessor sees visually most closely represents the volume of trash at a given assessment site. For this reason the BASMAA report recommends using the Qualitative Site Assessment Score as a means to reduce resources to obtain information on the amounts of trash at a site.

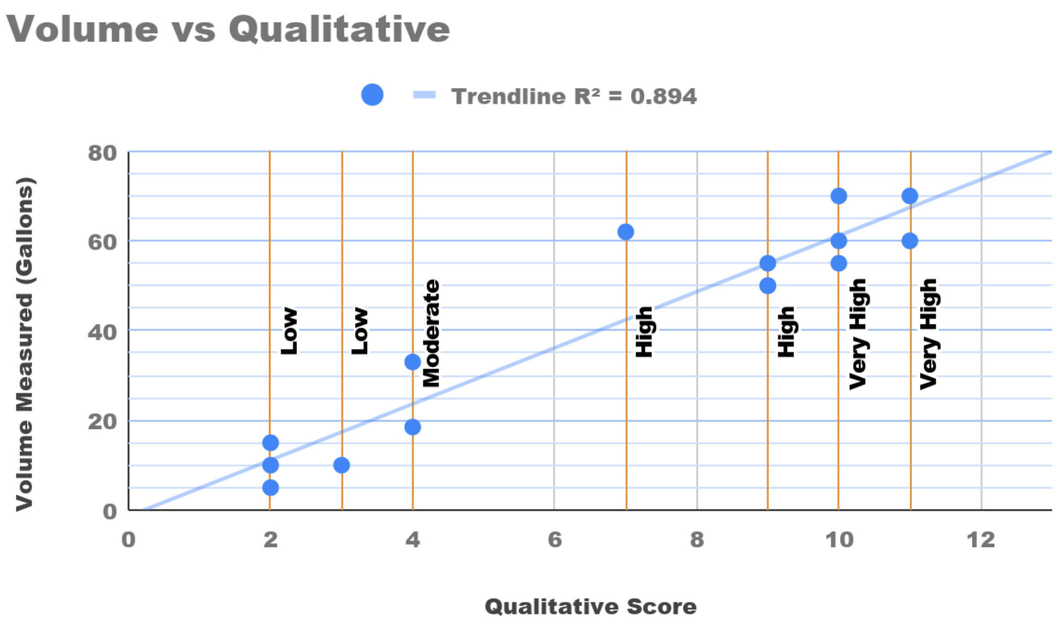


Figure 58. Qualitative Score versus Volume measured in gallons by site.

Additional studies took place during this study relative to methods tested for this project. BASMAA performed a pilot study of its methods in the Bay Area and a graduate student from Loyola University in Chicago, Illinois collaborated with the project team to perform some of these methods in Illinois. A brief summary of these additional studies can be found in the appendix of this document.

Acknowledgements

The authors of this document are indebted to many. We convened a Technical Advisory Committee that provided us valuable feedback. Members of the TAC included people who were already trash monitoring practitioners, machine learning/artificial intelligence experts, program-based method developers, testing experts, and funders. These folks were Chris Sommers of EOA, Donna Bodine of GeoSyntec (through the first year), Ted Von Bitner of Wood Environment & Infrastructure Solutions, Inc. (formally AMEC Foster Wheeler), Karine Wisenbaker of Aquatic Bioassay Consultants, Kate Wing of Kate Wing Consulting, George Leonard of the Ocean Conservancy, Sherry Lippiatt of NOAA, Kaitlyn Kalua of the California Coastkeeper Alliance, Holly Wyer of the Ocean Protection Council, and Greg Gearheart of the State Water Resources Control Board. Additional thanks go to Chris Sommers, his team at EOA, and BASMAA members for coordinating with the Method Evaluation Project Team in the Bay area. More thanks go to Karin Wisenbaker and SMC members for coordinating with the Method Evaluation Project Team in Southern California.

References

ASTM International. "Waste Management Standards." https://www.astm.org/Standards/waste-management-standards.html

Bay Area Stormwater Management Agencies Association (BASMAA). 2017. Receiving Water Trash Monitoring Program Plan for the San Francisco Bay Region. EOA, Inc. Version 1, October 2017.

Bay Area Stormwater Management Agencies Association (BASMAA). 2020. San Francisco Bay Area Receiving Water Trash Monitoring *Pilot-Testing of Qualitative and Quantitative Monitoring and Assessment Protocols.* EOA, Inc. Version 1, July 2020.

Calculator.net. "Standard Deviation Calculator." http://www.calculator.net/standard-deviation-calculator.html

Exchange Network Leadership Council. 2006. "Environmental Sampling, Analysis and Results Data Standards: Overview of Component Data Standards." https://www.epa.gov/sites/production/files/2015-06/documents/esaroverview_10012014a.pdf

Federal Aviation Administration. 2016. "FAA News." https://www.faa.gov/uas/media/Part_107_Summary.pdf

Moore, Shelly, Martha Sutula, Ted Von Bitner, Gwen Lattin, and Kenneth Schiff. 2016. "Southern California Bight 2013 Regional Monitoring Program: Volume III. Trash and Marine Debris." http://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/928_B13_Debris.pdf

Ode, P.R., A.E., Fetscher, and L.B. Busse. 2016. Standard Operating Procedures for the Collection of Field Data for Bioassessments of California Wadeable Streams: Benthic Macroinvertebrates, Algae, and Physical Habitat. California State Water Resources Control Board Surface Water Ambient Monitoring Program (SWAMP) Bioassessment SOP 004.

Rice University. "OnlineStatBook Project." http://onlinestatbook.com/2/power/factors.html

San Francisco Bay Regional Water Quality Control Board (San Francisco Bay Regional Water Board). 2004. Rapid Trash Assessment Protocol, Version 8.

http://www.swrcb.ca.gov/rwqcb2/water_issues/programs/stormwater/muni/mrp/WaterBoard%20Trash%20Assessment%20Method%20SWAMP_v8.pdf

Appendix - Additional Method Studies

During this Method Evaluation Study, other studies were taking place. The Project Team worked closely with these other projects. Here is a brief synopsis of the other projects and their results.

San Francisco Bay Area Receiving Water Trash Monitoring Pilot Testing Results

In October of 2017 the Bay Area Stormwater Management Agencies Association (BASMAA) submitted their Final Trash Monitoring Plan to the SF Bay Water Board staff (BASMAA 2017). Implementation of this Trash Monitoring Plan from October 2017 to July 2020 represented the "pilot-testing phase" of trash receiving water monitoring in the San Francisco Bay Area. During this time, the pilot protocols and methods were applied. Monitoring Plan objectives and scientific monitoring questions outlined in the Trash Monitoring Plan were used to guide the evaluation of trash monitoring and assessment data results presented in the final report entitled San Francisco Bay Area Receiving Water Trash Monitoring Pilot Testing Results (BASMAA 2020).

Two trash assessment methods were developed and used for the pilot testing phase of the Trash Monitoring Plan. Qualitative trash assessments were performed as visual surveys of trash levels (i.e., conditions) to derive a Site Assessment Condition Score. Trained personnel assign a trash condition score from 1 to 12 (12 being the most trash) to a site based on the level of trash that was observed both within the water body and along its banks or shoreline within a defined assessment area. The second method was a quantitative trash monitoring method that entailed removing, sorting and measuring the volume of trash found within an assessment area at a targeted site. Both the qualitative and quantitative assessment methods were used at targeted sites to allow for the comparison of both approaches.

A total of 125 urban creek, channel and riverine probabilistic sites throughout the MRP Area were qualitatively assessed for trash. A total of 625 qualitative trash assessments were conducted over five sampling events (three during wet season and two during dry season) between October 2017 and March 2020. A total of 100 targeted sites were selected for both qualitative and quantitative trash assessments. Additionally, a total of 200 trash assessments were conducted over two sampling events at targeted sites. Targeted monitoring was conducted at nine trash boom locations in Alameda, Santa Clara and San Mateo Counties.

Key Findings of the San Francisco Bay Area Receiving Water Trash Monitoring PIlot Testing Results include:

1. Significant correlations were observed between qualitative trash condition scores and trash density (volume per unit area) at both regional and countywide scale. The visual assessment tool was recommended as a valid approach to assess conditions when using volume of trash as the indicator for trash conditions.

2. Region-wide, approximately 77% of the urban stream lengths in the MRP Area exhibit low to moderate levels of trash.

3. Trash condition scores at targeted sites were generally higher (more trash), compared to probabilistic sites.

4. Seasonality appears to have no effect on trash levels observed/measured at receiving water sites. Trash levels were highly similar between the dry and wet seasons. Storm intensity and frequency did not appear to have an influence on trash levels observed during the wet season.

5. Litter/Wind and Other/Stormwater trash pathways were the most frequent pathways reported at all monitoring sites, however, Illegal Encampments and Illegal Dumping trash pathways were associated with the largest proportion of trash observed.

Both of these methods were also used as part of the Method Evaluation Study outlined in this report. Similarities include a high correlation between the qualitative site assessment score and the volume measurements.

Chicago Area Method Comparisons

During this project, the Project Team was introduced to a graduate student from Loyola University in Chicago, Illinois who was conducting trash assessments in rivers. To show compatibility of some of the methods assessed here in California with another state, Lauren Wisbrock, a graduate student working under Dr. Tim Hollein, from Chicago's Loyola University, performed two of the assessment methods included in this study. She included the qualitative visual assessment method and the quantitative tally method in streams and rivers near Chicago (Figure 59).

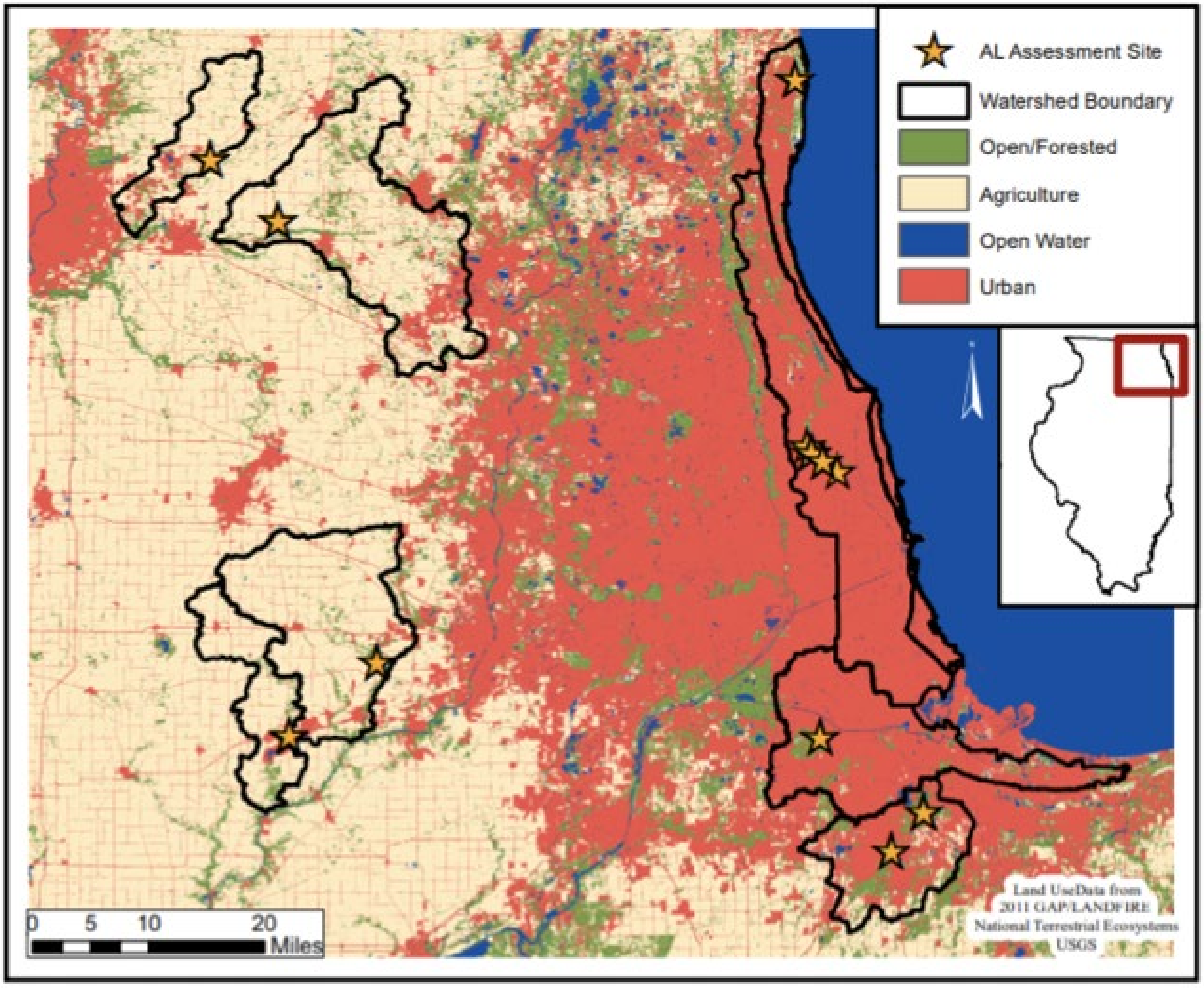


Figure 59. Map of Sampling Sites in and near Chicago, Illinois.

Initial results for her surveys are included here; however, it should be noted that she will be producing a more comprehensive report of her work, including these results, as part of the requirements for the completion of her degree.

Performance of the Tally Method was done two different ways for this study. Trash was tallied visually by walking the area and categorizing and counting the trash. The trash was then collected manually and categorized and counted back at the lab. The results were similar to the method evaluation study performed by the Project Team, showing that the amounts of trash counted in the field were less than that counted in the lab after the trash had been collected (Figure 24; see section on the Tally Method above).

While the numbers are much lower for the visual tally, there is a high correlation between the visual and manual methods (Figure 60), indicating that estimates could be made for the higher amounts of trash if necessary. However, caution should be used in doing this as there may be some breakage of items in the transport back to the lab. As with the method evaluation study, larger differences were seen in the plastic category where the breakage into smaller pieces has been observed and can account for much of the differences (Figure 61). Differences in the amounts glass for the Chicago Study were also large but may be an indication of breakage as well given the nature of glass.

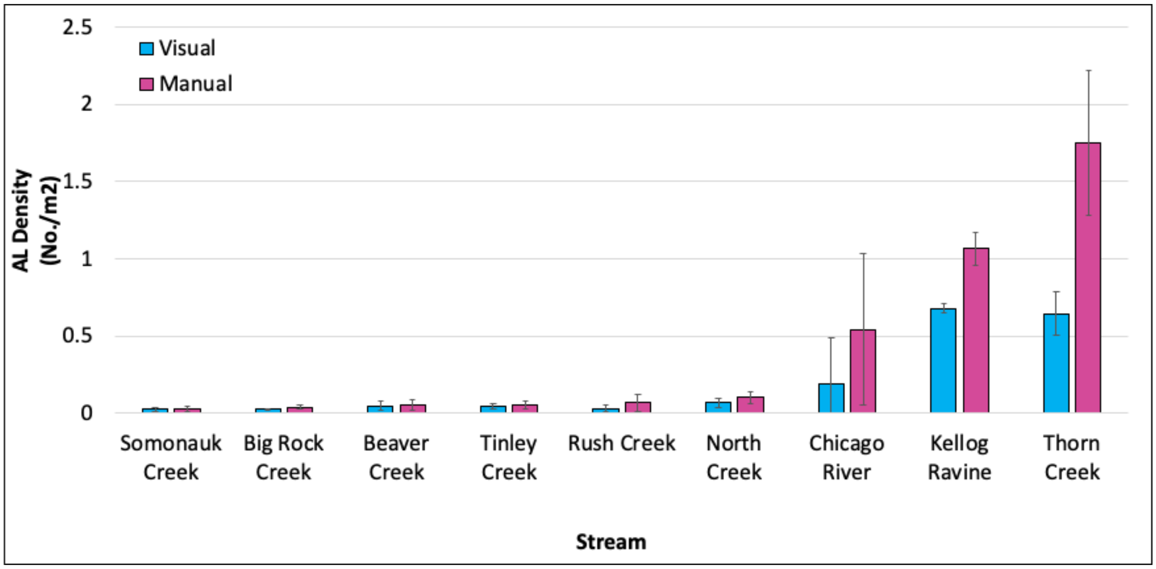


Figure 60. Amounts of Anthropogenic Litter (AL) at Sites in and near Chicago, Illinois.

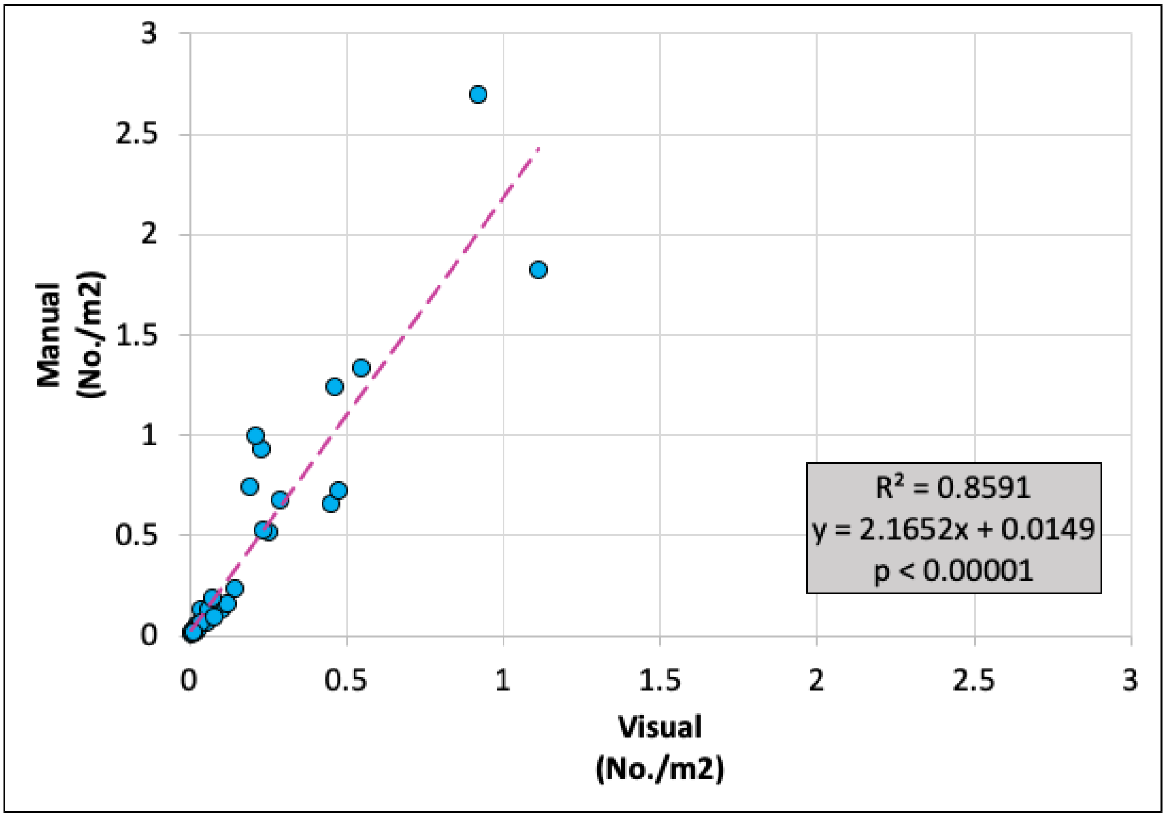


Figure 61. Correlation in the amounts of Anthropogenic Litter collected using a visual tally versus manually collecting and counting the trash back at the lab.

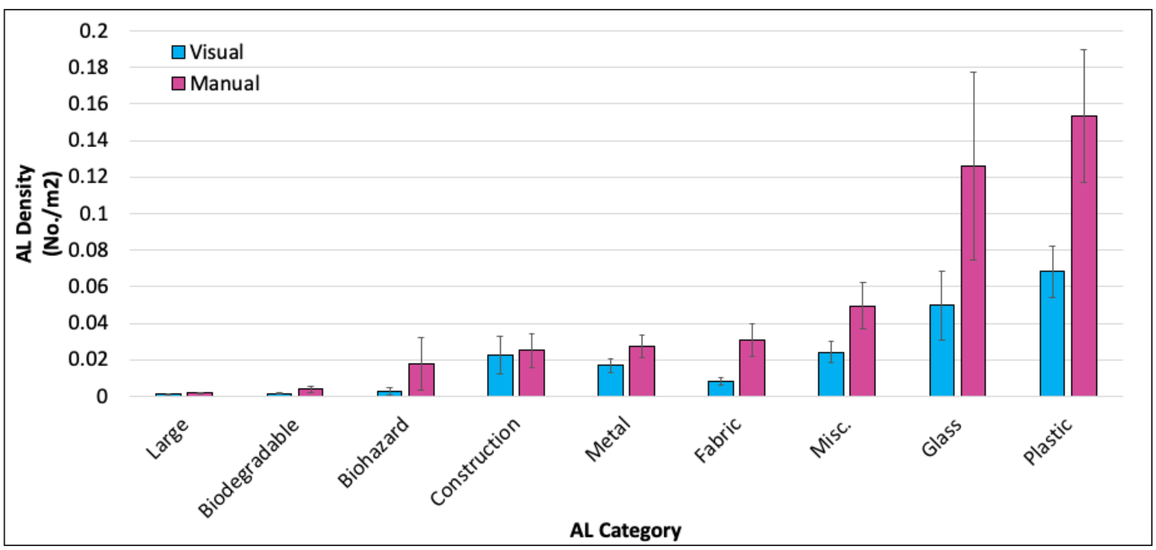


Figure 62. Correlation in the amounts of Anthropogenic Litter collected using a visual tally versus manually collecting and counting the trash back at the lab.

The Qualitative Method was also performed at the Chicago sites and comparisons between the manual count of trash and the qualitative assessment show a low correlation ($R^2 = 0.4705$; Figure 42). This is similar to the results for the same in the Method Evaluation study ($R^2 = 0.473$), indicating that the Tally Method and the Qualitative Assessment Scores are not good predictors for one another.

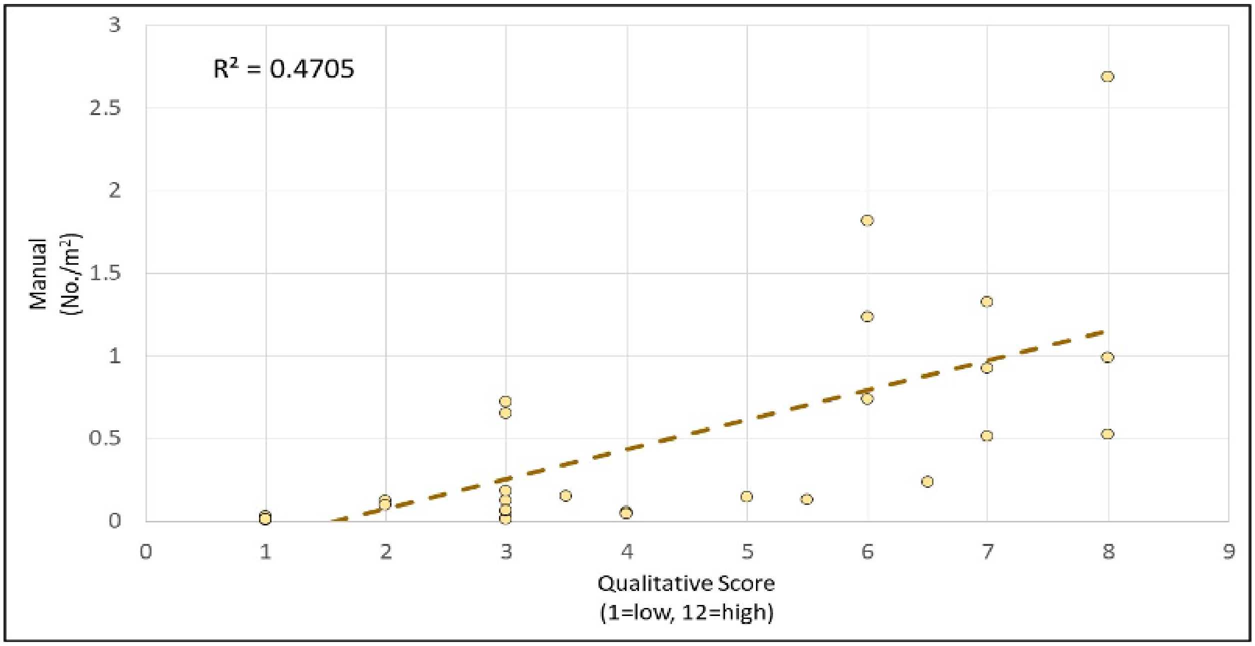


Figure 63. Manual Tally vs Qualitative Score

The performance of the Tally Method and the Qualitative Assessment Score Method in Chicago area rivers and streams appears to be comparable to those methods performed in California. While no comparison was made between trash items identified, the methods themselves were viable and presented similar results to this Method Evaluation Study.