TETRA TECH
9444 Balboa Avenue, Suite 215
San Diego, CA 92123
Tel 858-268-5746 • Fax 858-268-5809

FINAL Technical Memorandum

To:	**Ms. Natalie DeFreitas, Storm Water Department, City of San Diego**	**Date:**	**March 6, 2013**
From:	**Ms. Yvana Hrovat, Tetra Tech Inc.**	**Subject:**	**Volume:Weight Correlation Study - Draft Technical Memo**
cc:	**Mr. Clem Brown, Storm Water Department, City of San Diego**	**Project:**	**Contract H084445, Task Order 51**

Background

Tetra Tech was contracted through Task Order (TO) 51: Catch Basin Cleaning Pilot Study Phase 2 to make recommendations to improve the City of San Diego's (City's) catch basin maintenance procedures and program. The overarching goal of TO 51 is to characterize catch basin cleaning criteria, methods and frequency in order to optimize pollutant load reduction per unit of effort expended. This Technical Memorandum summarizes the results of Subtask 6 of TO 51, which aims to derive a correlation between debris volume estimates and weight for the purpose of pollutant load estimates.

Current catch basin cleaning practices do not explicitly measure pollutant loads, rather debris volume is estimated visually by field crews. As part of TO 38, pollutant loads were coarsely characterized for four different land use areas throughout San Diego as a percentage of debris weight. In order to apply this pollutant load characterization to debris removal data throughout the City and account for annual pollutant load reduction resulting from the catch basin cleaning program, debris weights are needed.

The purpose of Subtask 6 was to conduct a field study and statistical analysis to establish relationships between crew-estimated debris volume (cu ft.) and actual weight of debris (lbs.) removed in manual clean-outs by field crews. With the volume to weight relationships established, total pollutant loads removed resulting from the catch basin cleaning program can be estimated. Specifically, the resulting volume to weight relationships will be used to:

1) estimate annual pollutant load removal by watershed resulting from the catch basin cleaning program between 2008-2012 and,
2) if improved or more frequent cleanings are recommended as part of TO 51, project the pollutant removal benefit of increased cleanings.

Methods

Ron's Maintenance, Inc. (Ron's Maintenance) is contracted to clean every catch basin on public property (within the right of way and easements) within the jurisdiction of the City of San Diego Storm Water Division once each year, as required by the City's municipal permit (Order No. R9-2007-01). The crews collect data for each catch basin inlet cleaned that includes the basin number, the percentages each of sediment, plastic, and detritus, and the volume of material removed. Under Subtask 6, field crews shadowed Ron's crew and recorded both weight and volume of catch basin cleanouts for three varying land use regions throughout the City. Over the course of eleven days, MOE and Ron's Maintenance inspected 715 catch basins in the northern portions of the City of San Diego. Neighborhoods visited included Carmel Valley/Del Mar Heights, Mira Mesa, and Scripps Ranch. Of the 715 catch basins, 361 had no identified materials within the basins. In the remaining 354 basins, the average amount of material weighed 28.5 pounds per catch basin and was roughly 5 gallons in volume. Field methods are covered in further detail in Appendix A, Section 1. Field data was then statistically analyzed to determine volume-weight relationships that consider several influencing factors. Factors that were considered include debris composition, debris moisture content, and predominant land use surrounding the catch basins. Statistical analyses were performed to determine if these factors influenced the weight and volume of catch basin material. Since the correlation suggested a strong relationship between weight and volume, a regression analysis was performed to investigate predictability of weight and volume by each factor and sources of variability. Given that strong relationships existed, predictive equations were developed to calculate weight from volume. . Statistical analysis methods are covered in further detail in Appendix A, Section 2.

Results

Statistical analyses resulted in five predictive equations that can be used to estimate weight of material removed for previous and future catch basin clean-out events. These equations are shown in Appendix A, Table 3 and discussed in further detail in Appendix A, Section 3. The application of these equations requires the volume of material removed from an individual catch basin to be measured and the material to be categorized by type and moisture (where available). Three of the five equations can be used to estimate weight where moisture content is unknown. The equations should not be applied to samples with less than 0.5 gallons of material collected.

Statistical analyses suggest that weight can be predicted from volume using the derived equations under similar field conditions with 68-84 percent accuracy. When using these equations and considering debris removed on an annual sum basis, the differences between predicted and observed total loads removed from catch basin cleaning were calculated to range from 3.6 to 6.6 percent with 90 percent confidence. It is important to note that, due to bias at the high endpoints in the dataset, total weight determined by using these predictive equations is always slightly overestimated.

For the overall purposes of TO 51, loads removed from catch basin cleaning will be determined on an annual basis. The predictive equations that do not account for moisture will be used to estimate loads from previous catch basin clean-out data (2008-2012). These annual loads will be mapped and used to further refine catch basin cleaning prioritization criteria and provide recommendations on higher priority clean-out areas.

Additionally, the predictive equations will assist Comprehensive Load Reduction Plan (CLRP) Phase II modeling efforts by allowing for prediction of pollutant load reduction attributed to catch basin cleaning. Pollutant loads will be determined by multiplying average pollutant concentrations by annual weight removed by watershed. These loads will be subtracted form CLRP Phase II modeling results to account for pollutant load reduction attributed to the Catch Basin Cleaning Program. The detailed steps involved in determining pollutant loads are outlined in the Subtasks 2 and 4 Technical Memorandum.

Considerations

The accuracy outlined above is only applicable to data collection that is done under similar conditions and using the same methods and categorization that was performed for this study. Conditions of this study that should be considered in future applications of the predictive equations include the seasonal timing of the study and occurrence of rain fall events during field work. Because field work was conducted during winter months, the relationships derived tend to be reflective of more organic matter than usual as fallen leaves from deciduous trees were typically present throughout the study. Additionally, the several rain events that occurred during the period of field work resulted in more moist and saturated samples.

For future efforts, it should be considered that catch basin clean-out inspection logs include moisture content of material removed (recommended as part of Subtask 5 of TO 51) and the number of preceding days since the last a rainfall event. These efforts would help classify future catch basin clean-out events based on moisture content in addition to composition. With more physical characteristics of the material removed known, the predictive equations that consider moisture developed in this study can be used to estimate weights more accurately. Additionally, a year-round study of weight and volume relationships should be considered to develop regression equations that are reflective of each season or an equation that can be used for any time of year.

Appendices

APPENDIX A: FIELD METHODS AND STATISTICAL ANALYSIS

1 Field Methods

Mikhail Ogawa Engineering (MOE), as a sub-contractor to Tetra-Tech, performed the field work associated with Subtask 6. MOE accompanied Ron's Maintenance as they performed storm water catch basin inspection and clean-out services for the City of San Diego. If materials are present in the catch basins upon inspection, Ron's Maintenance collects the material in plastic bags and cleans the base of the catch basin. Crews from Ron's Maintenance clean each catch basin manually using a shovel, broom, and dust pan and collect debris in 5-gallon buckets. If no material is present, Ron's Maintenance notes that the catch basin is clean and continues to the next catch basin (MOE, 2013).

For the purposes of Subtask 6, MOE accompanied Ron's Maintenance to perform the field work within select neighborhoods in the northern portions of City of San Diego. Neighborhoods were chosen based on the predominant land use present and number of available catch basins to visit. Visiting catch basins in differing neighborhoods allowed analyses to consider land uses as a determining factor in weight and volume relationships. The land uses of focus for Subtask 6 included commercial/industrial use, residential use, and mix use and are illustrated in Figure 1. Detailed illustrations of the visited areas and contributing drainage areas are shown in Figure 2, Figure 3, and Figure 4, respectively. The neighborhoods visited for these land uses include Carmel Valley/Del Mar Heights, Mira Mesa, and Scripps Ranch.

This space intentionally left blank

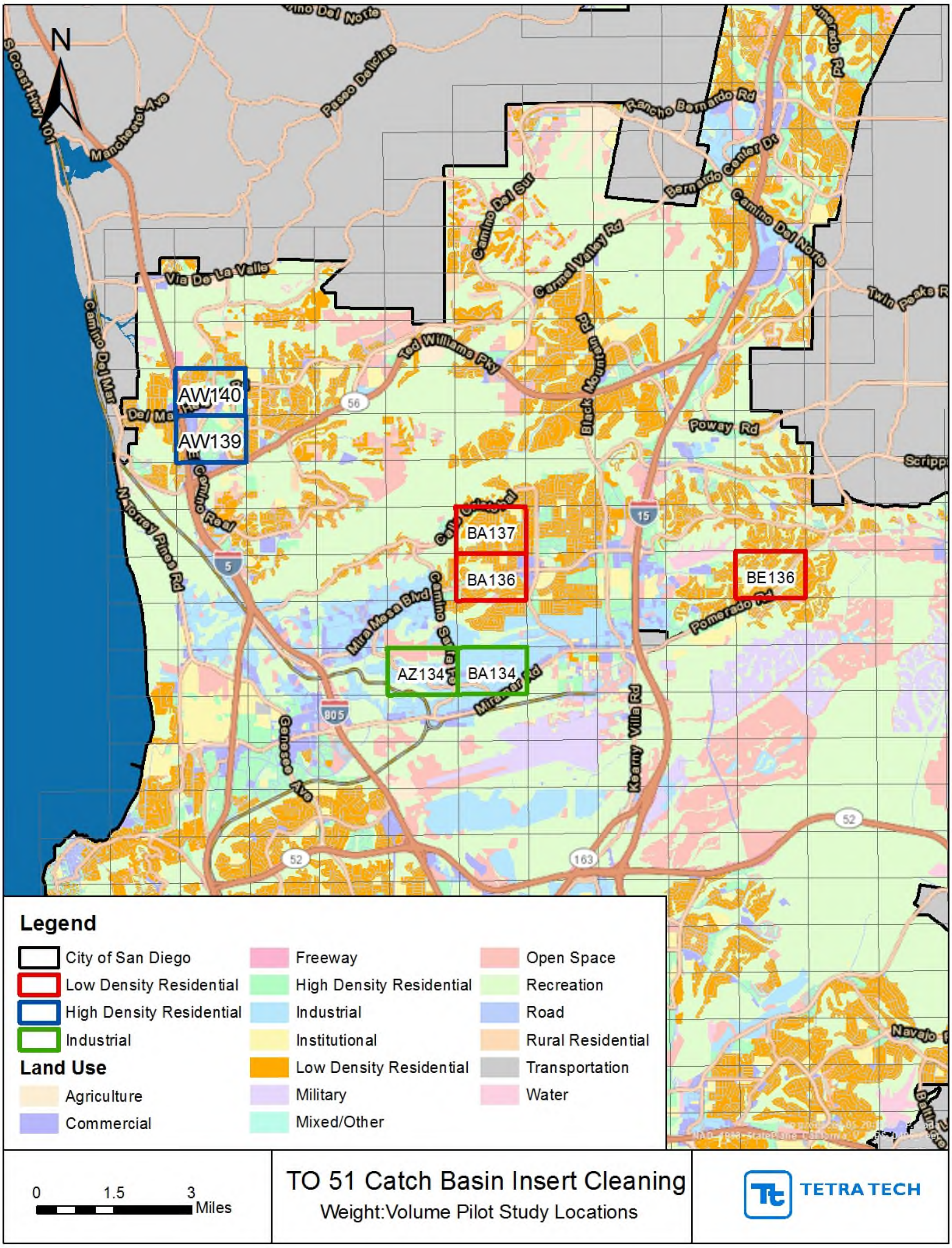

Figure 1. Volume:weight correlation study locations

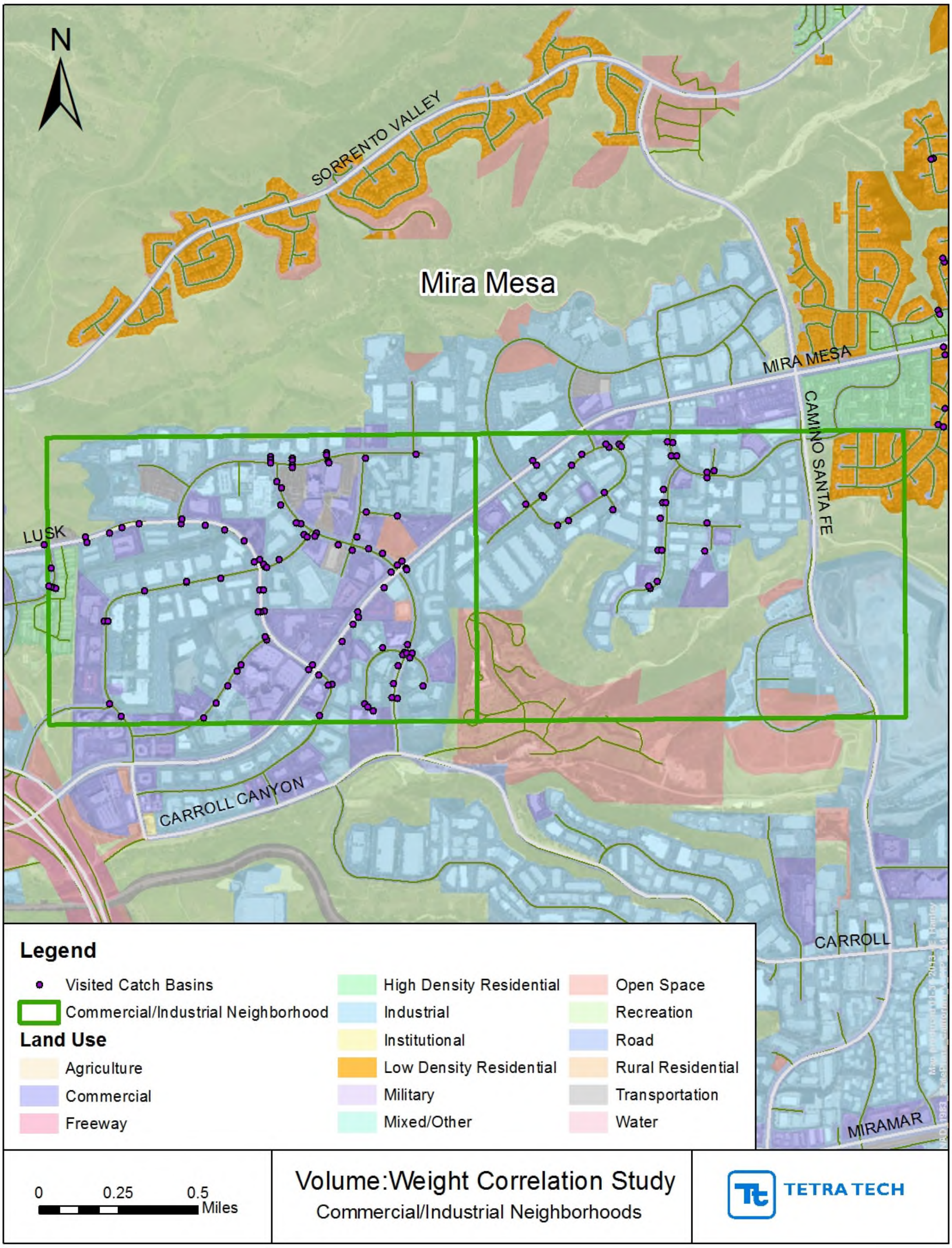

Figure 2. Commercial/Industrial land use neighborhoods

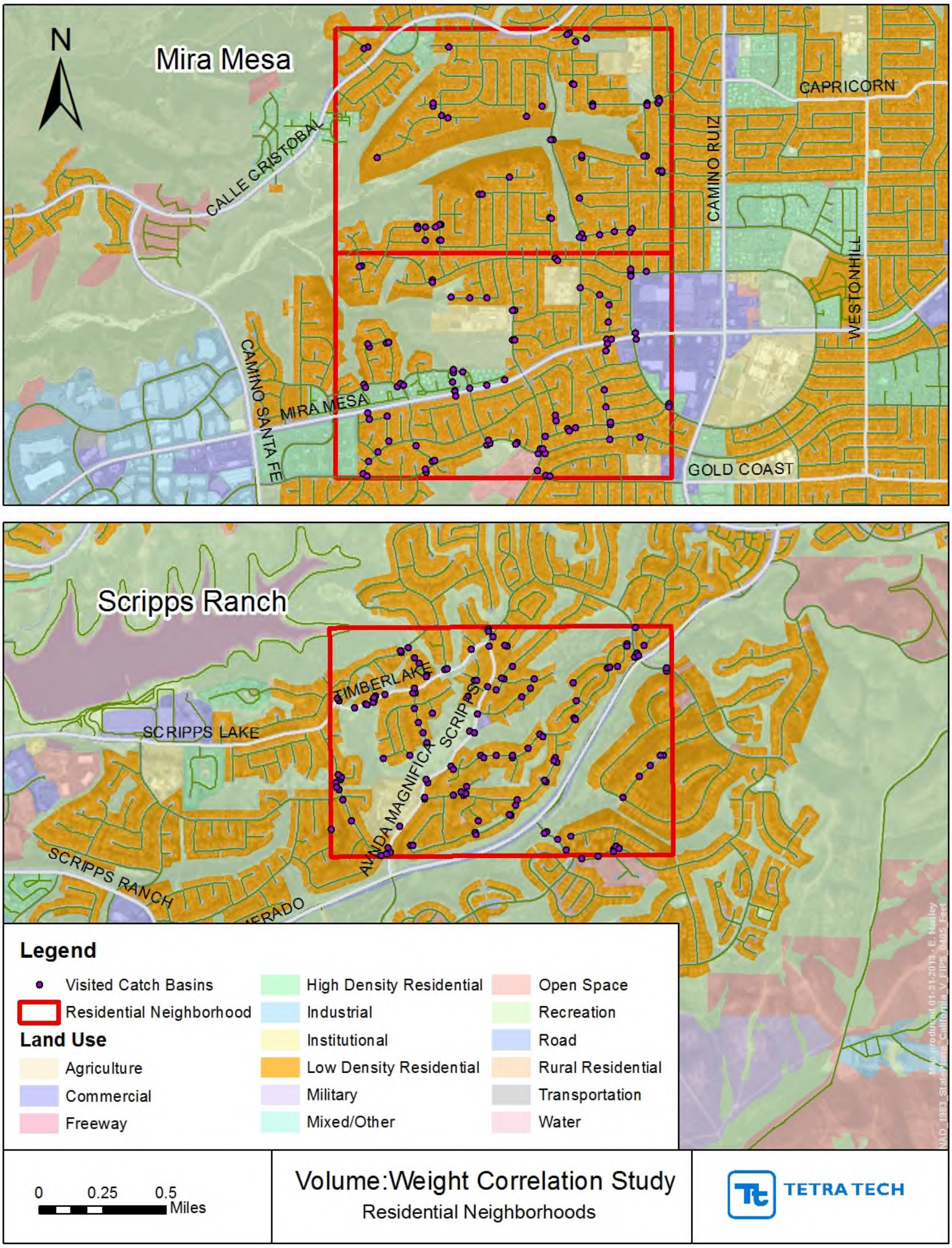

Figure 3. Residential land use neighborhoods

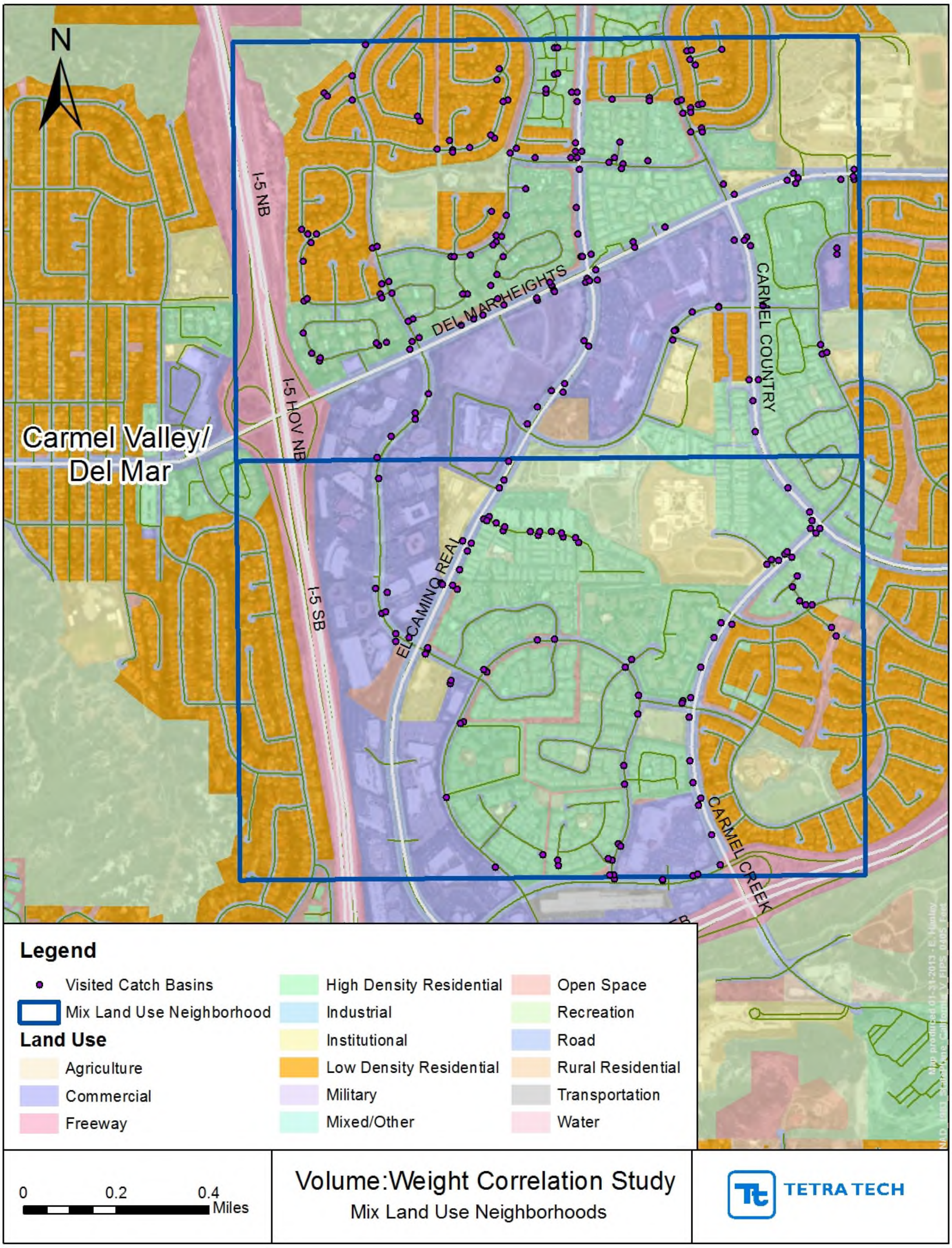

Figure 4. Mix land use neighborhoods

As Ron's Maintenance performed their contractual services, MOE collected information about the catch basin and the materials collected. Data collected by MOE included Basin ID, percent predominant land use, weight of material, estimated volume, moisture content of materials, and percent of estimated composition. Classifications used for indicating predominant land use, moisture content, and composition were as follows:

Land Use Categories:

- Single family residential
- Multi-family residential
- Commercial
- Industrial
- Parks/Recreation
- Open space/undeveloped

Moisture Content Categories:

- Dry
- Moist
- Saturated

Composition Categories:

- Sediment (mineral)
- Gravel/Rocks (mineral)
- Leafy (organic)
- Woody (organic)
- Plastic (litter)
- Paper/Cardboard (litter)
- Cigarette (litter)
- Metal (litter)
- Glass (litter)
- Polystyrene foam (litter)

Land use types draining to each catch basin were based on mapping and visual observations. Weighing of the material involved using a Chatillon Digital Hanging Scale and a steel can with swing handle. Volume was estimated in the field on a half-gallon basis and contents were inspected visually using best professional judgment. Moisture content was also determined by visual observation.

Over the course of eleven days, MOE and Ron's Maintenance inspected 715 catch basins in the northern portions of the City of San Diego. Neighborhoods visited included Carmel Valley/Del Mar Heights, Mira Mesa, and Scripps Ranch. Of the 715 catch basins, 361 had no identified materials within the basins. In the remaining 354 basins, the average amount of material weighed 28.5 pounds per catch basin and was roughly 5 gallons in volume. In residential areas, the composition of the materials generally included sediments and leafy material, which is consistent with the surrounding tributary areas. In commercial/industrial areas, the composition of materials included sediments, leafy materials and a higher percentage of litter (than residential areas), e.g., paper, plastics and polystyrene foam (MOE, 2013).

2 Statistical Analysis

Statistical analyses of field data acquired by MOE were performed to determine relationships between weight and volume of catch basin debris. A two-fold statistical analysis was performed involving a correlation analysis followed by a regression analysis. If strong correlations between weight and volume are evident, regression equations can then be developed to represent volume-weight relationships empirically. Regression analyses aim to develop a regression model that considers all the factors that affect the volume:weight relationship of catch basin debris. If a single regression model cannot be developed to represent all the factors using the entire dataset, than a series of regression models are developed for sub-datasets that represent specific factors (i.e., debris by land use, debris by composition, debris by moisture, etc.).

For this study, the correlation analysis was conducted in several ways with variation in the data subsets, transformations, and correlation technique. Since the correlation suggested a strong relationship between weight and volume, a regression analysis was performed to investigate predictability and sources of variability. The regression analysis pursued in this study was a general linear model (GLM) with categorical factors. The categorical factors considered include material types (sediment, organic, or mixed), material moisture (dry, moist, or saturated), and predominant land use surrounding the catch basins (low intensity residential, commercial/industrial, or mixed uses). Combination of these sub-datasets simplifies the application of weight:volume relationships as it reduces the overall number of regression equations in the end.

2.1 CORRELATION ANALYSIS METHODS

Correlations show whether two variables are related and to what degree. Application of correlation methods depends on whether or not the data is normally distributed. Pearson correlation is a parametric method that depends on normalized data distributions. Spearman correlation, on the other hand, is a non-parametric correlation of the ranks of values that does not require normal distributions. Both correlations were used in this statistical analysis as they provide interpretations of the dataset regardless of whether the condition of normality is met. Pearson and Spearman correlations coefficients indicate the strength of correlations present within the dataset. A correlation coefficient of 1 indicates a strong correlation; whereas, a correlation coefficient of 0 indicates no correlation.

2.2 REGRESSION ANALYSIS METHODS

If a strong correlation is evident through the correlation analysis, a regression analysis can be used to develop weight and volume regressions. For this study, several regression analyses were performed to categorically address influencing factors such as material composition, moisture content, and predominant land use surrounding the catch basin. To consider these factors independently and then conjointly, subsets of the data were used. For debris composition, subsets were categorized as sediment (≥70% of material was sediment or gravel), organic (≥70% of materials were leaves or woody), or mixed (when neither organic or sediment categories exceeded 70% of the sample). For moisture content, subsets were classified as saturated, moist or dry as noted in the field notes. Within this dataset, sample sizes were small (<20 samples) for certain combinations of moisture and material types, including dry sediment, saturated mixed, and saturated organic (Figure 5). For these subsets, grouping with other larger subsets was considered when practical.

This space intentionally left blank

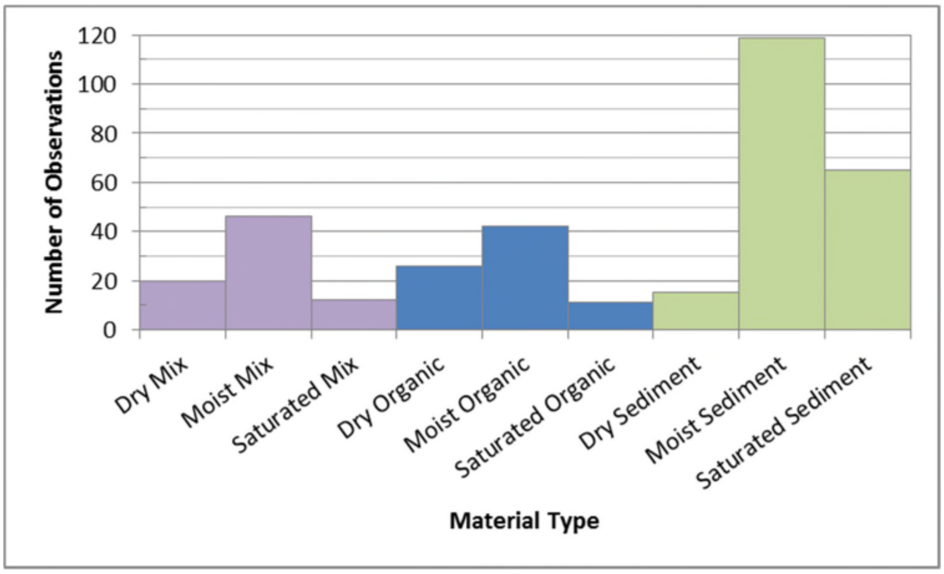

Figure 5. Numbers of observations in categories of debris composition (sediment, mixed, and organic) and moisture content (saturated, moist, and dry).

To analyze the subsets of data, a GLM was used to examine significant differences in regression slopes and intercepts between the data subsets. The factors that resulted in similar slopes and intercepts were grouped for developing general linear regressions that were more precise than a regression model of all data combined. The resulting regression models can be used for predicting weight from volume. For all regression analyses, the regressions were not forced through the origin. Therefore, application of the predictive equations should be limited to data points that are within the range of the calibration data (volumes >0.5 gallons).

3 Results

In the catch basin clean out study, roughly half of the records showed clean sumps, where catch basins weight equaled zero and volume equaled zero. These records are useful in estimating material weight and volume overall, but are not useful in establishing a relationship. For regression analysis and analyses with subsets by material type or moisture, clean sumps catch basins were excluded. There were 355 observations of clean catch basins.

3.1 Correlation Analysis Results

Both weight and volume distributions were skewed towards low values, especially when clean catch basins were included. Logarithmic transformation of weight and volume reduced skenwness and gave the best approximation of a normal distribution (Figure 5**Error! Reference source not found.**). Pearson coefficients indicate the strength of linear or log-transformed dependence between two variables, with higher numbers indicating greater strength, up to a maximum value of 1. Spearman's rank correlation coefficients give a nonparametric measure of statistical dependence of two variables. The Spearman correlation increases in magnitude (to a maximum of 1) as the two variables become closer to being perfect monotone functions of each other. Correlation coefficients for log transformed data and non-parametric methods were greater than 0.80 for all data subsets (Table 1), suggesting that a strong relationship between weight and volume exists overall and in certain groups. Correlations were strongest for weight and volume in the sediment dominated and moist subsets, which were also the largest subsets.

Note that an outlier with a volume greater than 100 gallons is not presented in Figure 5 but is included in the analyses.

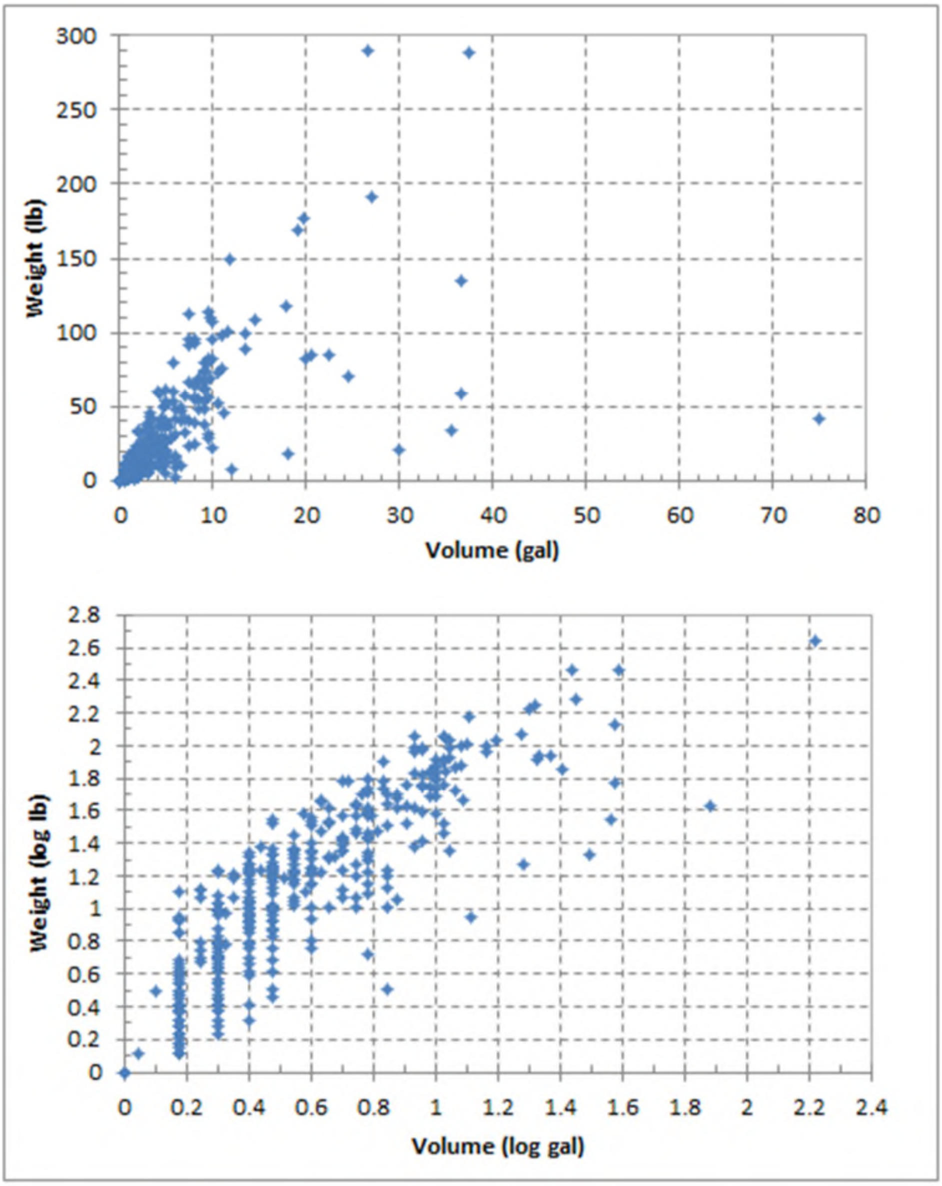

Figure 6. Relationships between weight and volume of catch basin materials without (above) and with (below) logarithmic transformation of data.

Table 1. Correlation coefficients for volume:weight relationships

Data subset	Sample size (N)	Correlation coefficients Pearson Untransformed	Pearson Log transformed	Spearman Rank
All valid readings	712	0.78	0.94	0.98
All non-zero readings	357	0.75	0.84	0.88
Sediment dominated	204	0.95	0.91	0.93
Organic dominated	78	0.54	0.81	0.84
Mixed material	74	0.93	0.89	0.93
Moisture: Dry	61	0.43	0.81	0.88
Moisture: Moist	207	0.83	0.87	0.91
Moisture: Saturated	89	0.87	0.84	0.85

3.2 Regression Analysis Results

Since a strong correlation was evident through the correlation analysis and especially amongst log-transformed data, a three factor GLM considering debris composition, moisture content, and predominant land use was used. Evaluation of the land use factors demonstrated that contributing land use had the least influence on weight to volume relationships. As shown in Figure 7, residential and commercial/industrial land use regression lines had similar slopes and intercepts. This similarity illustrates no statistical distinction between these land uses and their effect on the weight to volume relationship. Slope and intercept were only significantly different for the mixed land use category. Since distinct data trends cannot be drawn from each land use category, land use was dropped and further analyses considered only debris composition and moisture content as factors.

This space intentionally left blank

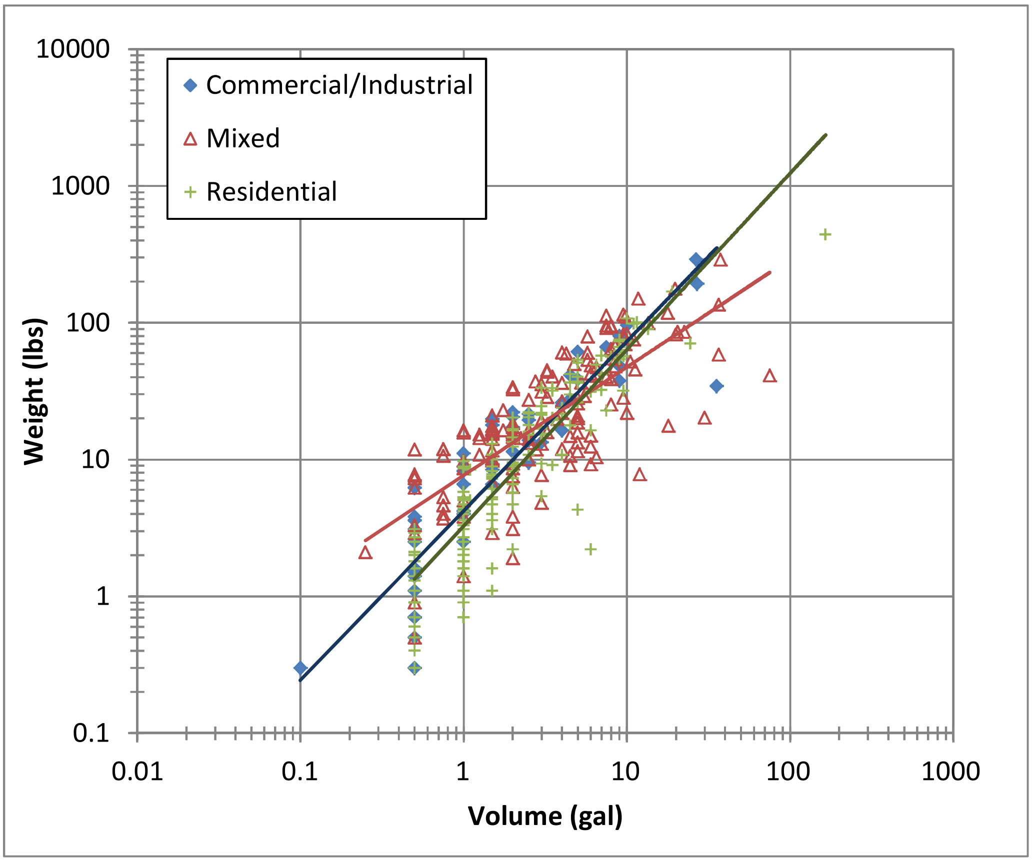

Figure 7. Relationship between volume and weight grouped by dominant surrounding land use types (commercial/industrial, mixed, and residential).

Since the land use factor was removed from analyses, a model which included volume, debris composition, moisture content and all possible interactions (a three way interaction term) was performed. As no significant three-way interaction model could be concluded, a two way model was then pursued. In a two-way model, the remaining factors (debris composition and moisture) are compared in pairs using T-statistics. T-statistics are statistical measures used to identify the difference between the means of two populations or groups of data. A large T-statistic indicates a small probability that the means of those two groups are the same. A small T-statistic is indicative of two compared groups where the means of those two groups are the same. T-statistics are interpreted in absolute value.

The two-way model started with 9 sub-groups of data that present debris composition and moisture content combinations:

- Dry mix
- Moist mix
- Saturated mix
- Dry organic
- Moist organic
- Saturated organic
- Dry sediment
- Moist sediment
- Saturated sediment

For application of T-statistics, the dry sediment group was used as the reference group to which other groups were compared. T-statistics (t-value) and probability values (p) were used to compare the slope and intercept of all the groups to the reference dry sediment group. The detailed results of this analysis are presented in Appendix B.

Assessment of T-statistics between the groups and the reference group allowed subset groups to be combined when slopes and intercepts were similar, suggesting that they can be considered as one group statistically. Combining groups simplifies the application of regression equations while instilling confidence in these equations as sample sizes increase. Debris composition was determined to be the most significant factor differentiating categories. In considering moisture content, dry content samples and moist content samples displayed similarities in slope and intercept suggesting that these two categories can be combined statistically. Unlike the dry and moist content groups, saturated content had distinct slopes and intercepts to warrant its own regression model. This T-statistics analysis resulted in 6 subset groups as listed in Table 2.

Table 2. Regression subsets by debris composition and saturation

Material Type	Number of Samples
Unsaturated Sediment (dry and moist)	138
Unsaturated Mix (dry and moist)	63
Unsaturated Organic (dry and moist)	67
Saturated Sediment	66
Saturated Mix	11
Saturated Organic	11

The relatively small sample size of the saturated organic group (N = 11) and the saturated mix group (N=11), makes their predictability very limited when considered independently. To increase confidence in regression equations, the saturated organic group and saturated mix group were combined with the unsaturated organic group and mix groups, respectively. For these instances, there is more confidence in developing relationships for organic and mixed debris types collectively rather than using equations that consider their moisture content. After this combination, the resulting four groups included: 1) unsaturated sediment, 2) saturated sediment, 3) mixed debris, and 4) organic debris. The regression lines for these final four groups are presented in Figure 8.

This space intentionally left blank

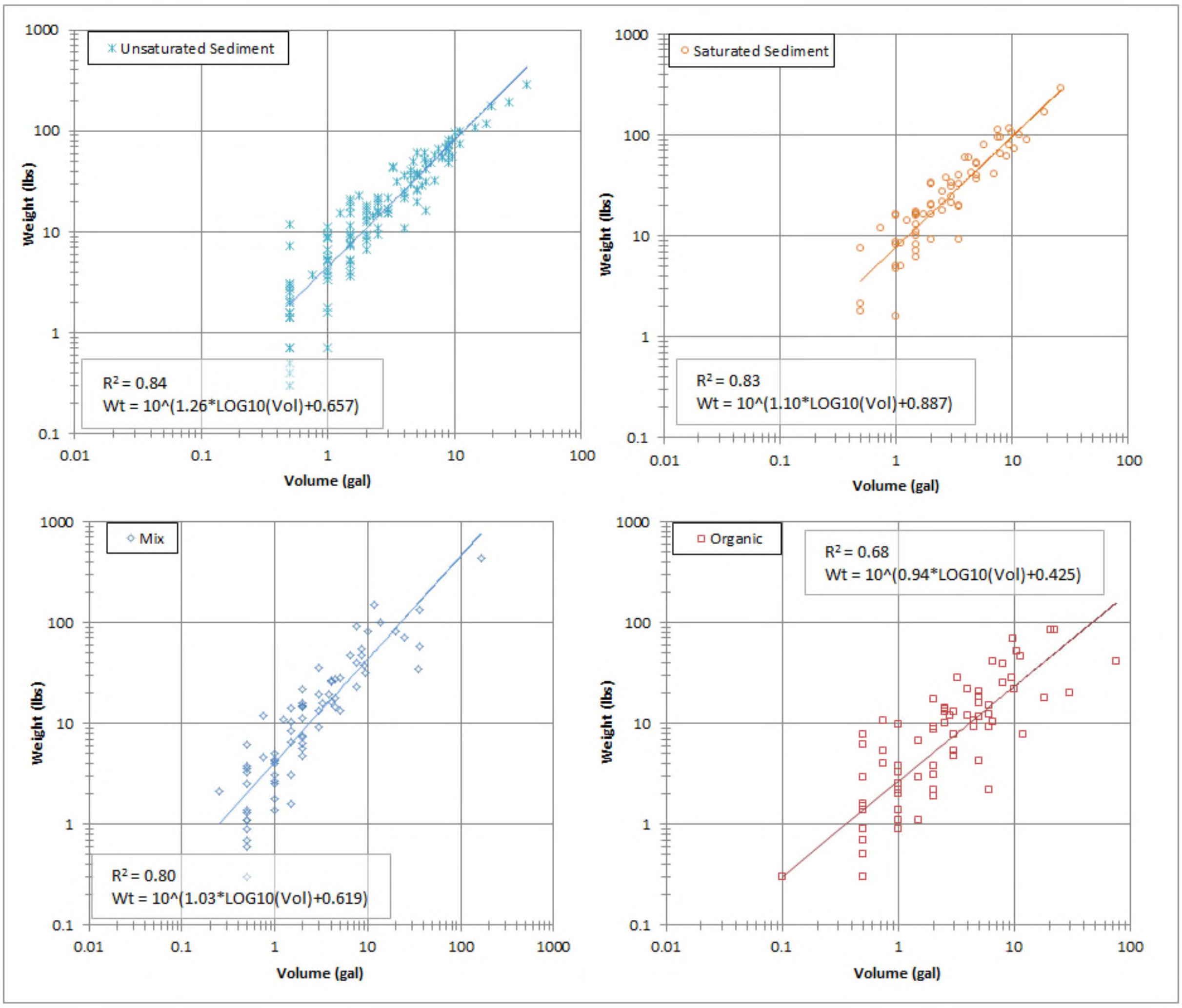

Figure 8. Regression relationships for four groups (unsaturated sediment, saturated sediment, mix, and organic)

Although the results of this subtask provides four confident regressions for four catch basin debris types, a regression for all sediment debris is useful in determining weight when moisture content is not available. The regression when considering sediment debris regardless of moisture content is presented in Figure 9.

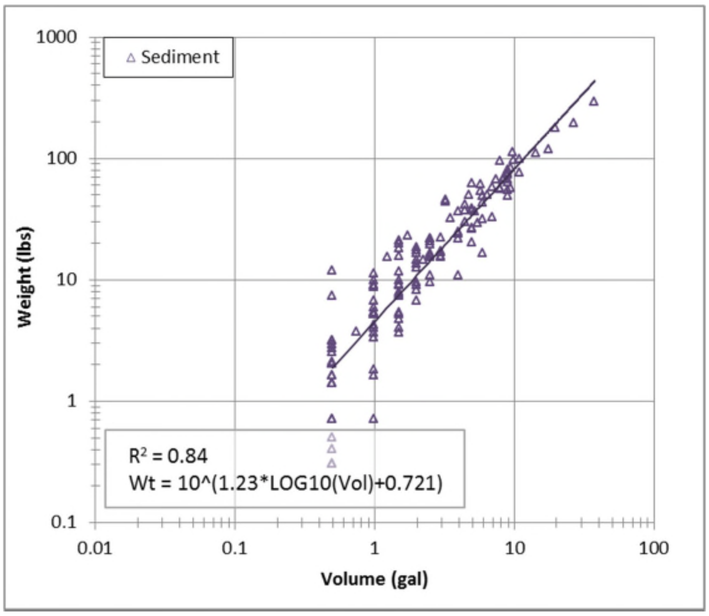

Figure 9. Regression relationship for sediment debris (all moisture contents)

Regression equations for the five debris categories from Figures 8 and 9 are presented in Table 3. The coefficient of determination term, R^2, is used to describe how well a regression line fits a set of data and provides a measure of how well future outcomes are likely to be predicted by that derived regression equation. It reflects the proportion of variability in a data set that is accounted for by the statistical model.

An R^2 near 1.0 indicates that the derived regression line fits the data well. As presented in Table 3, the R^2 terms for the five predictive equations are between 0.68 and 0.84, suggesting that weight can be predicted from volume using these equations under similar field conditions with 68-84% accuracy. These equations can be used to predict weight from volume in their respective material type and moisture categories.

When using these equations and considering debris removed on an annual sum basis, the differences between predicted and observed total loads removed from catch basin cleaning were calculated to range from 3.6 to 6.6 percent with 90 percent confidence. It is important to note that, due to the bias at the high endpoints in the dataset, total weight determined by using these predictive equations is always overestimated.

This space intentionally left blank

Table 3. Predictive equations for three categories of materials collected from catch basins.

Material type and moisture	Predictive equation	R^2
Unsaturated sediment (categorized as dry or moist and ≥70% sediment or gravel)	10^(1.26*LOG10(Volume_gal) + 0.657)	0.84
Saturated sediment (categorized as saturated and ≥70% sediment or gravel)	10^(1.10*LOG10(Volume_gal) + 0.887)	0.83
Mixed material (<70% sediment and gravel and <70% wood and leaves, all moisture types)*	10^(1.03*LOG10(Volume_gal) + 0.619)	0.80
Organic (≥70% of materials are leaves or woody, all moisture types)*	10^(0.94*LOG10(Volume_gal) + 0.425)	0.68
Sediment (≥70% of materials are sediment or gravel, all moisture types)*	10^(1.23*LOG10(Volume_gal)+0.721)	0.84

* These classifications and predictive equations are recommended to be used when moisture content of the material removed is unavailable.

4 References

Mikhail Ogawa Engineering (MOE). 2013. Memorandum. City of San Diego Task Order 51: Catch Basin Cleaning Study Phase 2. January 24, 2013.

APPENDIX B: REDUCED LINEAR REGRESSION MODEL RESULTS

Line	Model components		Values		
1	Formula: Weight as a factor of volume, moisture, material type, and interactions of volume with type and moisture		log(Weight_lb) ~ log(Volume_gal) + Moisture + MatrlType + log(Volume_gal):Moisture + log(Volume_gal):MatrlType,		
2	Residuals for the reduced model		Min -2.08, first quartile -0.33, median 0.07, third quartile 0.36, max 1.85		
3	Residual standard error		0.6263 on 346 degrees of freedom		
4	Multiple R-squared and adjusted R-squared		0.953 and 0.9516, respectively		
5	F-statistic		700.9 on 10 and 346 degrees of freedom		
6	p-value		< 2.2e-16		
7	Model terms (compared to the dry sediment group)				
8	# Coefficients:	Estimate	Std. Error	t value	p
9	log(Volume_gal) (Slope)	1.26	0.09	14.06	< 0.001
10	Intercept	1.57	0.12	13.17	< 0.001
11	Moisture: Moist	-0.08	0.11	-0.66	0.51
12	Moisture: Saturated	0.49	0.14	3.42	< 0.001
13	Material Type: Mixed	-0.15	0.11	-1.38	0.17
14	Material Type: Organic	-0.61	0.11	-5.81	< 0.001
15	Slopes				
16	log(Volume_gal):MoistureMoist	0	0.09	-0.39	0.97
17	log(Volume_gal):MoistureSaturated	-0.18	0.11	-1.67	0.096
18	log(Volume_gal):MatrlTypeMixed	-0.22	0.07	-3.09	0.002
19	log(Volume_gal):MatrlTypeOrganic	-0.31	0.08	-4.08	< 0.001

APPENDIX C: FIELD DATA SHEETS