Restoring and Protecting Our Waterways: Bioassessment Tools and Priorities for Action

Supplemental Environmental Project – Summary Report

June 7, 2021

Presented To

City of San Diego Transportation & Storm Water Department
9370 Chesapeake Drive, Suite 100
San Diego, CA 92123

Presented By

Tetra Tech, Inc.
9444 Balboa Avenue
Suite 215
San Diego, CA 92123

1-858-268-5746
1-858-609-1625
clint.boschen@tetratech.com
www.tetratech.com

The City of
SAN DIEGO

TETRA TECH

Contents

Acronyms/Abbreviations

Acronym/Abbreviation	Definition
BCG	Biological Condition Gradient
CADDIS	Causal Analysis/Diagnosis Decision Information System
CalFire	California's Department of Forestry and Fire Protection, Fire and Resource Assessment Program
CASTool	Causal Assessment Screening Tool
City	City of San Diego
MHPA	Multi-Habitat Planning Area
MSCP	Multiple Species Conservation Program
MWMP	Municipal Waterways Maintenance Plan
NASVI	Native Aquatic Species Vulnerability Index
NHD	National Hydrography Dataset
SCA	Screening Causal Assessment
SMC	Stormwater Monitoring Coalition
SEP	Supplemental Environmental Project
RPP	Restoration and Protection Prioritization
RPPTool	Restoration and Protection Prioritization Tool
RPS	Recovery Potential Screening
SCAPE	Stream Classification and Priority Explorer
SCCWRP	Southern California Coastal Water Research Project
WAMP	Watershed Asset Management Plan
Water Board	San Diego Regional Water Quality Control Board
WMP	Watershed Management Plan
WoE	Weight of Evidence
WQIP	Water Quality Improvement Plan

1.0 Introduction

The overarching goal of this Supplemental Environmental Project (SEP) was to provide an advanced, user-friendly toolkit that was designed to identify the likely causes of current biological impairment in streams within the City of San Diego (City), as well as the Southern California Stormwater Monitoring Coalition (SMC) region, and to prioritize stream reaches for future restoration and protection actions. Four major objectives were addressed in this SEP, as summarized from the Work Plan submitted to the San Diego Regional Water Quality Control Board (Water Board):

1. Develop an improved Screening Causal Assessment (SCA) methodology by incorporating a validated comparator site approach, additional diagnostics, taxa traits, better use of available site data, and more informative findings.

2. Incorporate the latest research on flow ecology into the SCA approach. Update and refine the flow ecology metrics and results based on more detailed genus/species level data and statistical analyses. Identify recommendations for future flow ecology studies/enhancements.

3. Develop a restoration and protection prioritization (RPP) framework, incorporating the concepts of recovery potential to help the City of San Diego determine high priority sites for restoration or protection based on the main stressors present and the potential for ecological lift.

4. Integrate the SCA approach into the RPP framework to create a SCA-RPP tool to increase the utility of the results for decision-making and appropriate allocation of City resources for the most beneficial activities.

To address these objectives, the City consulted with Tetra Tech to lead development of two companion bioassessment tools in association with the Water Board and the Southern California Coastal Water Research Project (SCCWRP). The Causal Assessment Screening Tool (CASTool, referred to as a Screening Causal Assessment (SCA) methodology in the SEP Work Plan), provides information indicating which stressor(s) are likely causes of biological impairment at a site. The Restoration and Protection Prioritization Tool (RPPTool), provides a screening application for decision makers to assign priorities to stream segments for restoration or protection. Together, CASTool and RPPTool comprise a bioassessment toolkit that all water resource managers in the SMC region, including the Water Board, can use to screen for likely stressors and prioritize sites for restoration or protection. Figure 1 schematically shows the bioassessment toolkit and relationships between the tools.

CASTool incorporates an innovative comparator site approach, biological diagnostics, and better use of available site biological data. It also incorporates the most current flow ecology (altered hydrology stressor) metrics for the SMC region. RPPTool incorporates stressors identified as likely causes of biological impairment based on CASTool, along with analyses of recovery potential factors (factors that most contribute to ecological uplift), threats, and co-benefit opportunities in a systematic way to prioritize streams for restoration or protection.

A Shiny web application was developed for each tool to provide an automated, user-friendly interface to support decision makers within the San Diego Region. For CASTool, this web application allows users to screen sites for the likely causes of biological impairment or identify potential stressors to which currently unimpaired sites may be vulnerable in the future. For RPPTool, the Shiny web application provides a user-friendly platform for prioritizing sites for restoration or protection depending on the user's objectives.

The following briefly summarizes the bioassessment toolkit and examples showing how the toolkit can be used to prioritize sites for restoration or protection. Detailed information regarding CASTool and RPPTool, including how input data are processed and analyzed, is provided in separate appendices to this report.
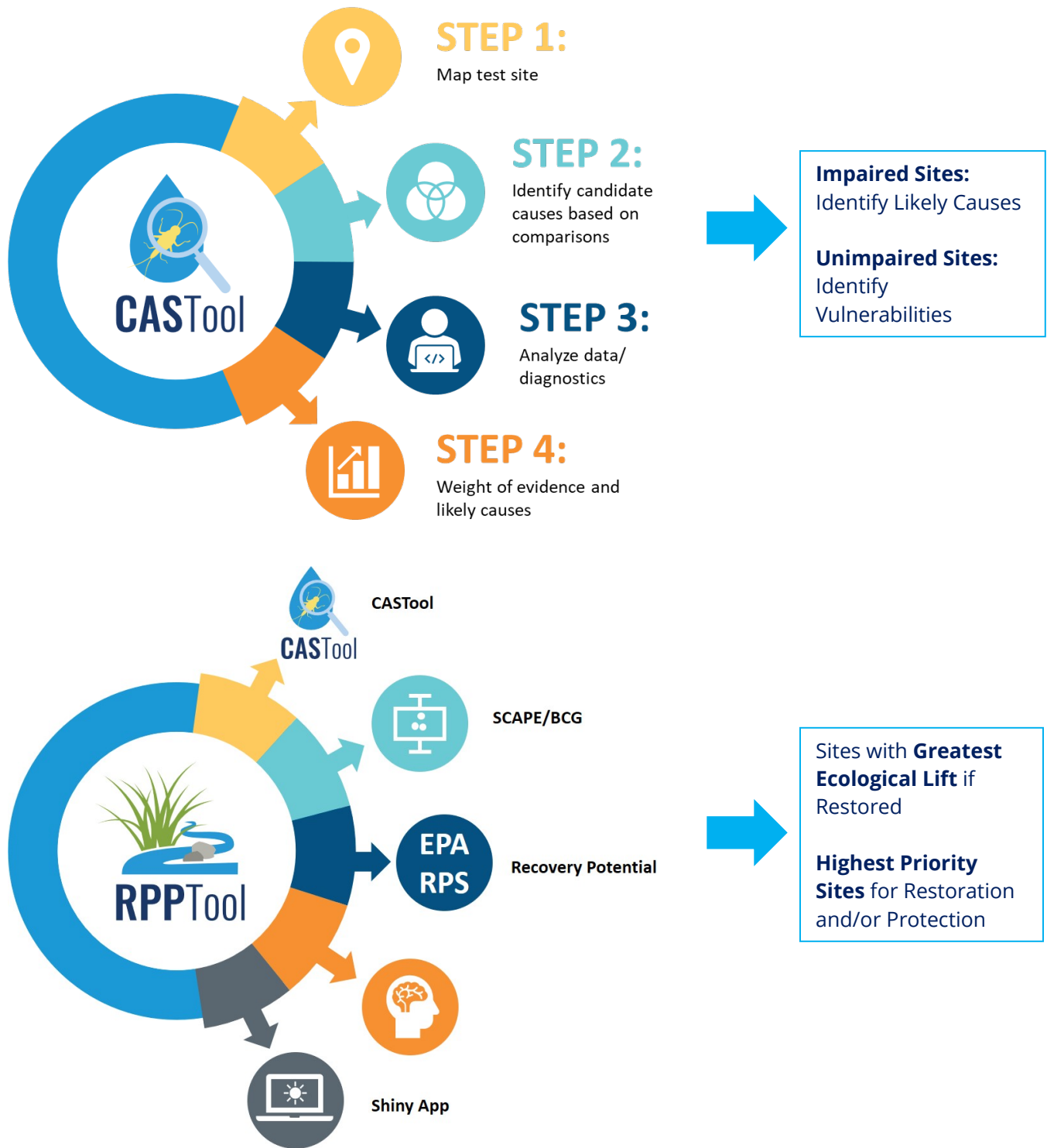
STEP 1:
Map test site

STEP 2:
Identify candidate causes based on comparisons

STEP 3:
Analyze data/ diagnostics

STEP 4:
Weight of evidence and likely causes

Impaired Sites:
Identify Likely Causes

Unimpaired Sites:
Identify Vulnerabilities

CASTool

SCAPE/BCG

EPA RPS   Recovery Potential

Shiny App

Sites with **Greatest Ecological Lift** if Restored

Highest Priority Sites for Restoration and/or Protection

Figure 1-1. Schematic diagram illustrating the relationship between CASTool and RPPTool and the types of inputs and outputs for each. Together, CASTool and RPPTool comprise a bioassessment toolkit that can address multiple objectives related to stream bioassessments.

2.0 Causal Assessment Screening Tool (CASTool)

CASTool uses an adaptation of the approach used in USEPA's Causal Analysis/Diagnosis Decision Information System (CADDIS, https://www.epa.gov/caddis), which focuses on the major lines of evidence (types of data and analyses) that are typically available based on routine bioassessments conducted in the SMC region. Comparator sites are used to assess biological conditions at sites of interest and to help identify either potential causes of biological impairment (if the site is impaired according to the proposed numeric Biological Objectives for the San Diego Region, based on the California Stream Condition Index [CSCI] score), or identify potential stressors to which the site may be vulnerable (if the site is currently unimpaired biologically).

Within the SMC region it is common to have multiple watershed scale non-point and point sources of potential stressors that affect biological conditions at a site depending on the extent of human activity and alterations to the landscape. To address this challenge, CASTool utilizes an innovative statistical approach developed by Tetra Tech to identify different groups of comparator sites based on non-biological factors that are known to affect distribution of benthic macroinvertebrate species and other aquatic life (e.g., elevation, slope, and underlying geology), as well as expected taxa derived from the statewide Observed/Expected (O/E) taxa model (Mazor et al 2016). These groups of comparator sites are then applied in analyses of site data.

CASTool includes a framework to streamline and automate selected analyses within USEPA's CADDIS. The automated analyses include the incorporation of altered flow metrics, stressor-specific diagnostic indicators including benthic macroinvertebrate tolerance values for specific water quality stressors (e.g., conductivity), and use of a transparent scoring system for summarizing which stressors are the most likely to affect the biological community at a site.

The general flow of analyses in CASTool is schematically shown in Figure 2. CASTool consists of multiple functions within a software package written in the Statistical Programming Language R that runs the analyses with little user input required other than the site(s) of interest. The package can be run through the CASTool web-based interface for most users, or for users with an understanding of the R language, from the R console. If running CASTool from the R console, multiple sites can be analyzed in one "batch". When using the web-based application, CASTool is run separately for each site. The R-console option will be made available on a case by case basis at this time.

Stressor data analyzed for each site in CASTool (if available) include:
- Water quality data (e.g., nutrients, metals, pesticides, other organic chemicals, and physicochemical parameters including dissolved oxygen, pH, conductivity, specific ions, dissolved solids)
- Stream habitat quality data (state-wide Index of Physical Integrity [IPI] and associated metrics)
- Altered flow metrics (flow ecology metrics and modeled data for the SMC region; Stein et al 2017)

Water quality data collected at various sites within the SMC region may be inconsistent due to different study purposes and to meet different requirements in some cases. Specific water quality parameters for which data are available may vary among sites, which is typical when integrating environmental data originating from different programs.

Biological response data analyzed in CASTool include:
- Benthic macroinvertebrate data (CSCI and component subindices and metrics)
- Algal data (ASCI and component metrics)[1]
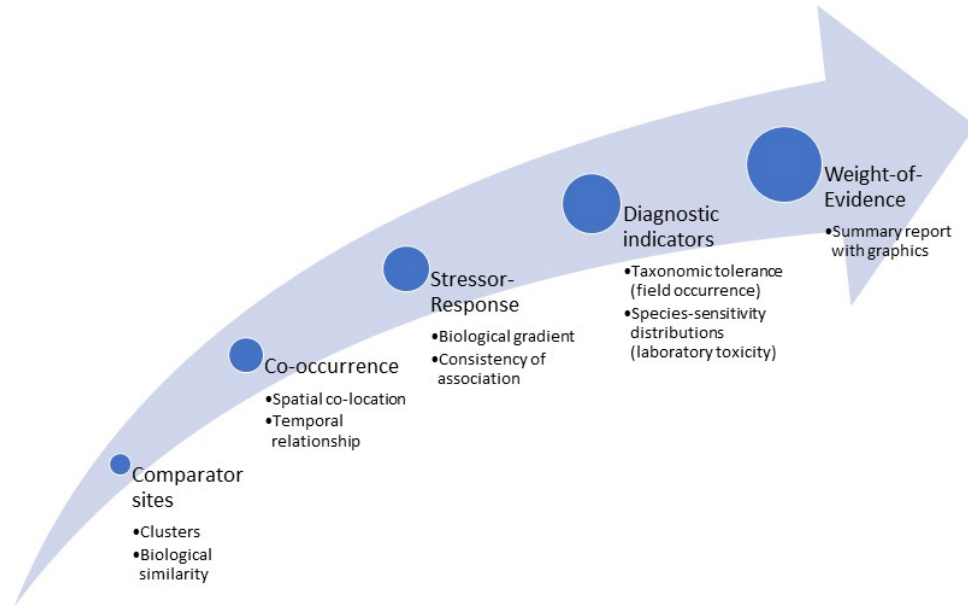
Figure 2-1. Graphic conceptual diagram of the modules in CASTool. Stressors include water quality, pollutants, toxics, altered flow, and habitat alterations. Responses include benthic macroinvertebrate and algae taxa characteristics.

CASTool evaluates specific lines of evidence that are frequently used in traditional causal assessments (e.g., CADDIS) and provide the strongest evidence as to whether a stressor is likely or not likely as a cause of concern at a site. The lines of evidence are:

- *Spatial/temporal co-occurrence* **-** Does the biological effect occur in the same location and at the same time as the stressor and not when or where the stressor is absent?
- *Stressor-response relationships from the field or elsewhere* **-** Does the adverse response increase or decrease as the stressor increases or decreases?
- *Temporal sequence* - Does the cause precede the adverse effect?
- *Verified prediction* – Does the prediction that taxa sensitive to a specified stressor occur less frequently at the impaired site than at sites with better biology?
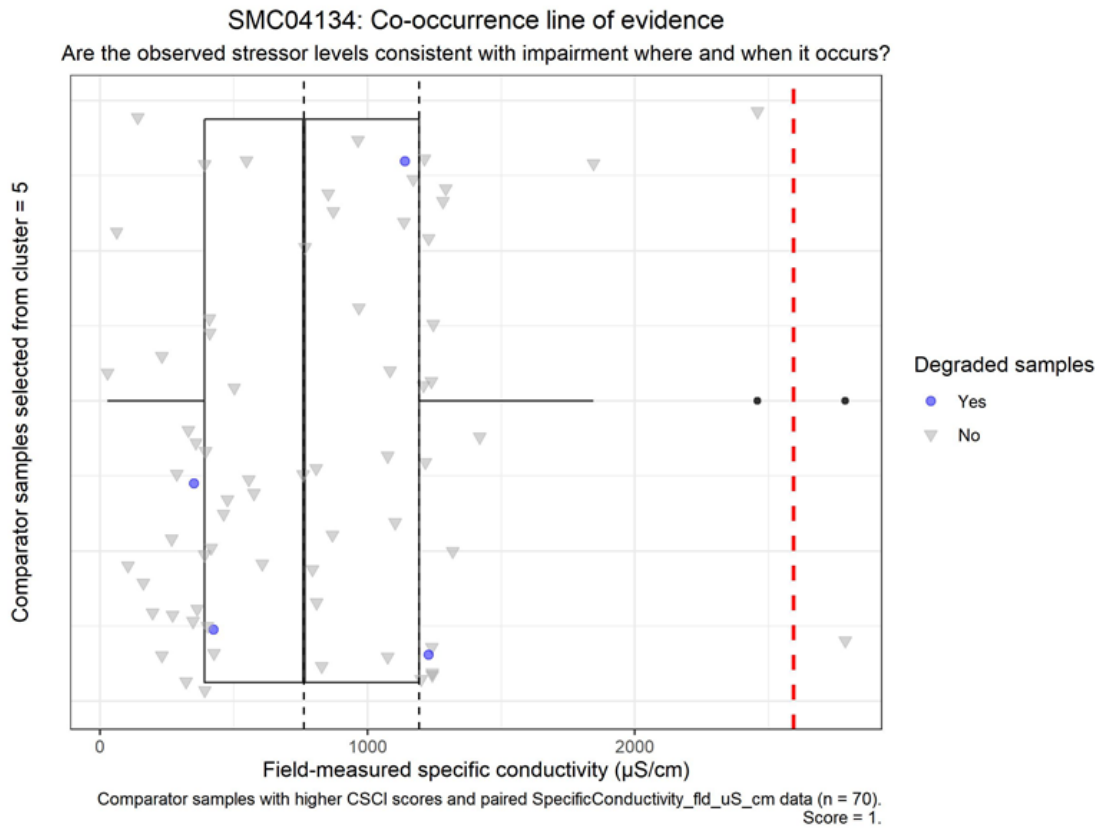
Graphical results of each of these lines of evidence, along with displays of the input data (stressor data and biological data) for the site are produced by CASTool. Figure 3 shows examples of two types of graphical output displayed by CASTool using results of different lines of evidence. Similar to USEPA's CADDIS framework, the results of analyses for each line of evidence are scored in terms of whether the results *support* the stressor as a cause of observed biological condition, *refute* the stressor as a cause, or are *ambiguous* (i.e., either conflicting or inconsistent results). In addition, transparent criteria are used within

---

[1] Note: The statewide algal index (ASCI) was published after the development of the tools. CASTool is designed so that it can incorporate the algal index in the future.

CASTool to determine whether results of a line of evidence strongly or weakly *support* a stressor, or strongly or weakly *refute* a stressor as a likely cause of observed biological condition.
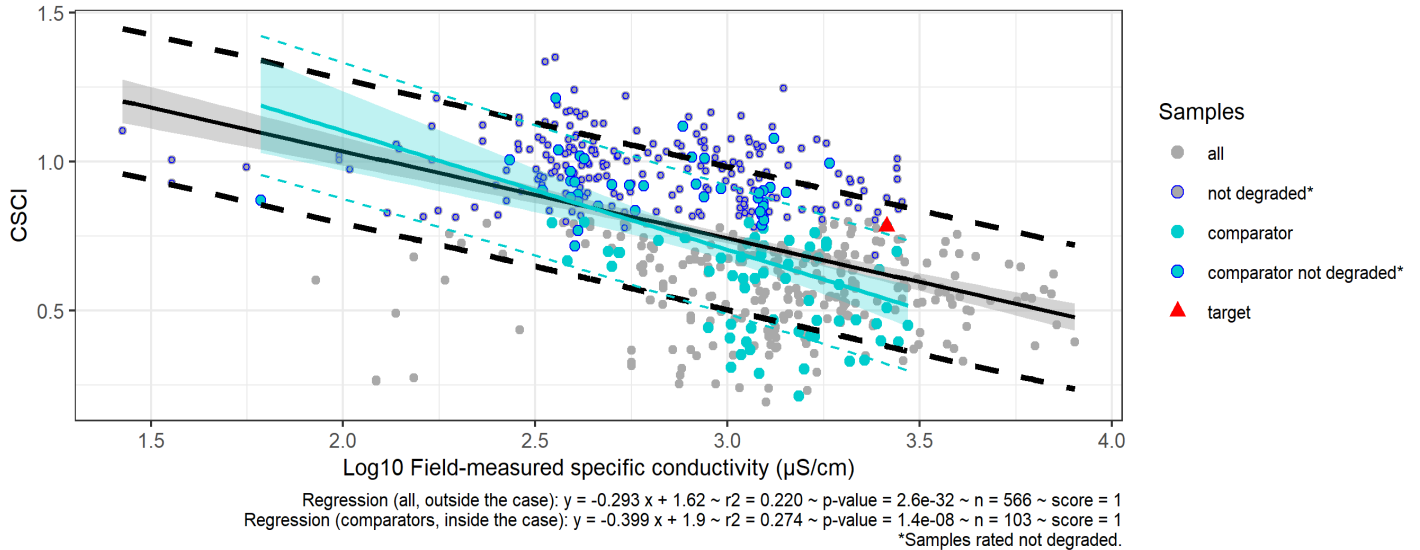
Scores from each line of evidence are incorporated in a Weight of Evidence (WoE) analysis similar to the approach used in USEPA's CADDIS. A WoE table is displayed showing how the scores were derived for each line of evidence and the overall score for each of the candidate causes (stressors) identified by CASTool for a site (see Figure 4 for an example).

SMC04134: Co-occurrence line of evidence
Are the observed stressor levels consistent with impairment where and when it occurs?

Comparator samples with higher CSCI scores and paired SpecificConductivity_fld_uS_cm data (n = 70).
Score = 1.

SMC04134: Stressor-Response (linear regression) line of evidence
Is there evidence of a biological gradient from inside or outside the case?
Linear regression with 75th percentile prediction interval

Regression (all, outside the case): y = -0.293 x + 1.62 ~ r2 = 0.220 ~ p-value = 2.6e-32 ~ n = 566 ~ score = 1
Regression (comparators, inside the case): y = -0.399 x + 1.9 ~ r2 = 0.274 ~ p-value = 1.4e-08 ~ n = 103 ~ score = 1
*Samples rated not degraded.

Figure 3. Examples illustrating two of the types of graphical results displayed for certain lines of evidence using CASTool

Weight of Evidence

Overall Findings

Summarized weight of evidence data are found in the file '911S01142_BMI_WoE_ExecSummary.tab' in the 'Results/911S01142/BMI/WoE' folder. More detailed weight of evidence data are found in the file '911S01142_BMI_WoE_ScoresTable.tab' in the 'Results/911S01142/BMI/WoE' folder.

Scores for each line of evidence and overall weight of evidence, weighted by number of stressors in each group. An overall weight of evidence score of 1 indicates that all evaluated lines of evidence for all observed stressors in the group support the group as a cause of impairment. An overall weight of evidence score of -1 indicates that all evaluated lines of evidence for all observed stressors in the group refute the group as a cause of impairment.

SiteID	Degraded	Group	Stressor_Samples	Stressors	Overall_WoE	TS_Inside	CO_boxplot	SR_Inside_LogRegr	SR_Inside_LinRegr	SR_Outside_LinRegr	VP_Boxplot
911S01142	No	Habitat	1	2	1.0	0	1.0	0.5	0	1.00	NA
		Ions	2	1	1.0	0	0.5	0.5	1	1.00	NA
		Metals_metalloids	1	1	1.0	0	1.0	1.0	0	1.00	NA
		Nutrients	2	1	0.0	0	1.0	0.0	0	0.00	NA
		Water quality	1	1	0.0	0	1.0	0.0	0	0.00	NA
		Modeled flow metrics	2	4	0.5	0	0.5	0.5	0	0.75	NA

Figure 4. Example of Weight of Evidence scoring results for a biologically impaired site using CASTool. In this example, nutrients and routine water quality parameters were identified as unlikely causes of impairment, whereas aspects of habitat quality and altered flow, as well as high concentrations of certain ions and metals are likely causes of impairment at this site.

CASTool has been applied to most of the bioassessment sites for which data are available as of 2020. Because there is a lag between the time biological sampling is conducted and when the data are available from the SMC data portal (as of this writing, the lag time is at least 2-3 years), the most recent data analyzed thus far are from 2017, with most of the currently available stressor and biological response data between 2008 and 2015 (Figure 5).
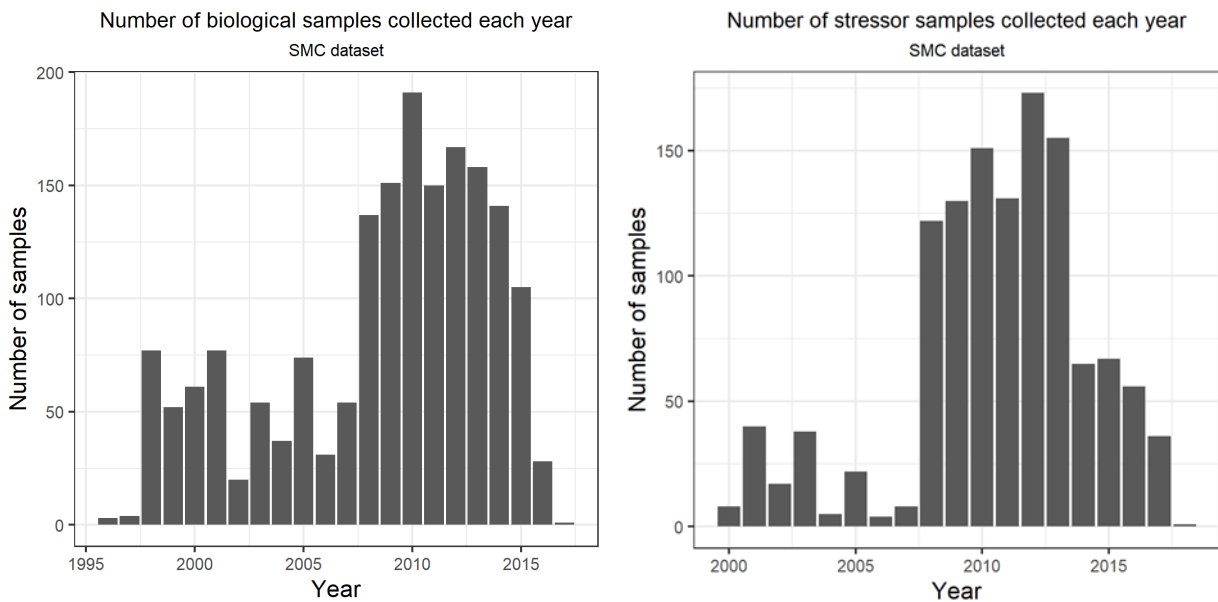
Figure 5. Number of biological and stressor samples in the SMC dataset by year between 2000 and 2020.

3.0 Restoration and Protection Prioritization Tool (RPPTool)

RPPTool produces an RPP Index score which is comprised of three subindices—Potential, Threats, and Opportunities—that combine to produce a restoration or protection index score for each reach depending on whether the site is impaired biologically (and therefore a candidate for restoration) or not (candidate for special protection). Figure 6 schematically illustrates the components used in RPPTool. The restoration or protection index scores are defined as follows:

- Restoration Index Score – the ecological uplift that could be attained through restoration of the reach, resulting in significant improvement of the biological community from one biological condition class to a higher class, incorporating likely threats and opportunities that would affect restoration success.

- Protection Index Score – the relatively high ecological condition of the reach that can be preserved through protection, that is, prevention of degradation of biological community from one biological condition class to a lower class, while accounting for threats and opportunities that would affect protection success.

Figure 6. Structure of RPP index with three subindices and nine component indicators.

Much like USEPA's Recovery Potential Screening (RPS) Tool, the results of Potential, Threats, and Opportunities subindices are scored individually and then combined to form the overall RPP index score. This tool, while similar to USEPA's RPS tool, focuses on local and regional California-specific data, as opposed to national data, and calculates indicators based on the catchment scale, as opposed to the watershed scale. RPPTool also incorporates bioassessment models and tools developed for California, such as the Stream Classification and Priority Explorer (SCAPE; Beck et al 2019), and the statewide Biological Condition Gradient (BCG) model developed by Tetra Tech in collaboration with bioassessment experts throughout California (Paul et. al. 2020). Regional data for current and predicted land uses and the Multiple Species Conservation Plans are also utilized by RPPTool. These data and models, along with the state-wide expected taxa model for benthic macroinvertebrates, also incorporated in RPPTool, provide a California-based restoration and protection prioritization approach with which to evaluate the potential for ecological uplift at a site.

In addition to using the latest models to characterize potential for ecological uplift, RPPTool also incorporates two indicators of potential threats to a site, fire and future population growth, as well as co-benefits and other opportunities which the user may want to include in the prioritization process. Non-ecological factors such as cost, feasibility, and site-accessibility for restoration or protection activities are not currently considered in RPPTool. However, use of RPPTool, combined with other tools used by the City to target management efforts, could provide a useful approach for achieving meaningful improvements in aquatic life.

The stream reach is the basic unit for evaluation in RPPTool. Stream reaches for the SMC region were obtained from the National Hydrography Dataset (NHD), version 2, NHDFlowline and NHDPlusCatchment shapefiles and the PlusFlowlineVAA table (USGS and USEPA, 2012).

Potential Subindex

Observed biological data and calculated metrics, including the final CSCI score, are obtained from the SMC data portal and used in CASTool. RPPTool uses the most recent observed CSCI score and stressor information to develop a weight of evidence to evaluate causality. The observed CSCI scores for all sites in the SMC region were obtained from the SMC data portal, whereas stressor data are available for sites that were analyzed using CASTool. Figure 5 displays the temporal range of sampling data in the entire dataset; observed stressor or response data may or may not be available for any given reach.

In RPPTool, the user may limit the temporal range of data used. The default range is present day through the past twelve years (currently 2008-2020). RPPTool also incorporates the distribution of predicted CSCI scores for each reach (where available) using landscape constraints as modeled in the SCAPE tool for the SMC region. Observed CSCI scores are compared to the reach's predicted CSCI distribution and corresponding BCG tiers from the state-wide BCG model to help determine potential for biological uplift at a site or reach (Figure 7).

RPPTool also incorporates a Biological Connectivity Indicator as part of the Potential subindex (Figure 6), which is a unique advancement for examining potential ecological uplift and ultimately restoration or protection priorities. The Biological Connectivity indicator represents the hydrologic connectivity of the target reach to upstream or (optionally) downstream reaches that may be in better biological condition (See Figure 8 as an example). Connectivity with nearby reaches having better biological condition than the target reach indicates the likelihood that source populations are nearby that may facilitate biological uplift from

restoration or protection. A higher connectivity score (due to better biological condition in nearby stream reaches) raises the biological condition score for the Potential subindex.
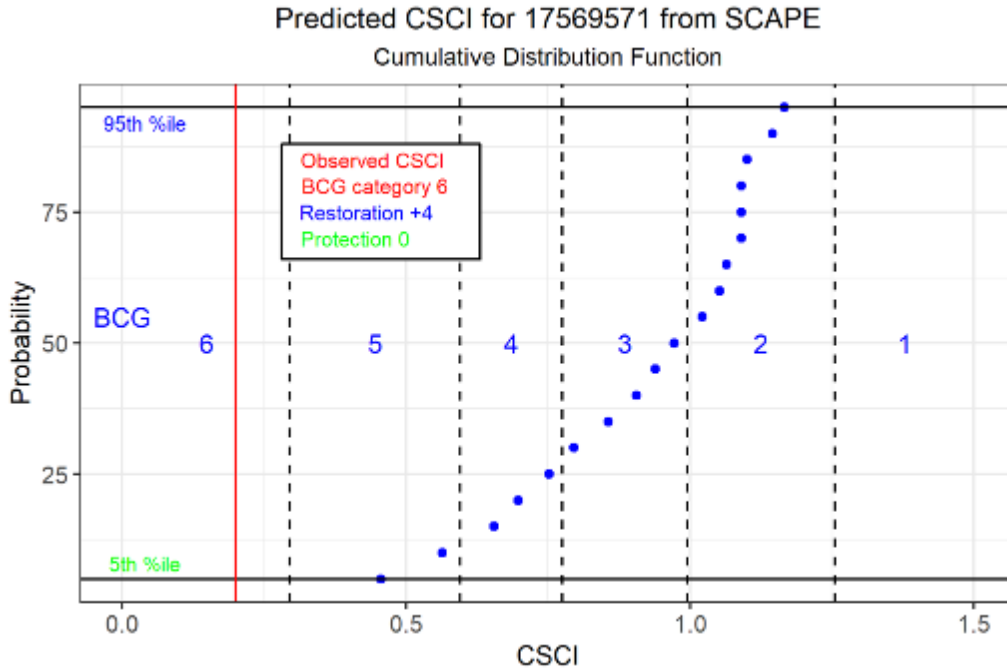
Figure 7. Example showing the modeled BCG tiers overlaying the SCAPE tool results for a site. The vertical red line represents the site's current biological score (CSCI and BCG tier). In this case, the site is capable of higher biological potential than it is currently attaining based on comparable sites in the SMC region.

Stressor information is obtained from CASTool described previously. The number of stressors present and the intensity of each stressor at a site are combined into a standard score that can depress the biological condition score in the Potential subindex calculations (Figure 6). The user can choose to weight each likely stressor derived from CASTool as 1 (default), 0 (not considered), or 2 (twice as important) to provide flexibility for user knowledge and experience with particular stressors in the region and to account for a given stressor likely being consistent with local background conditions (0), or resulting from a non-point source which would be more difficult to mitigate (2). Akin to the Biological Connectivity Indicator mentioned above, RPPTool calculates a Stressor Connectivity Indicator if appropriate stressor data are available for nearby upstream reaches. Downstream reaches are not considered in the Stressor Connectivity Indicator because most water quality stressors (e.g., chemicals, temperature) move from upstream to downstream.
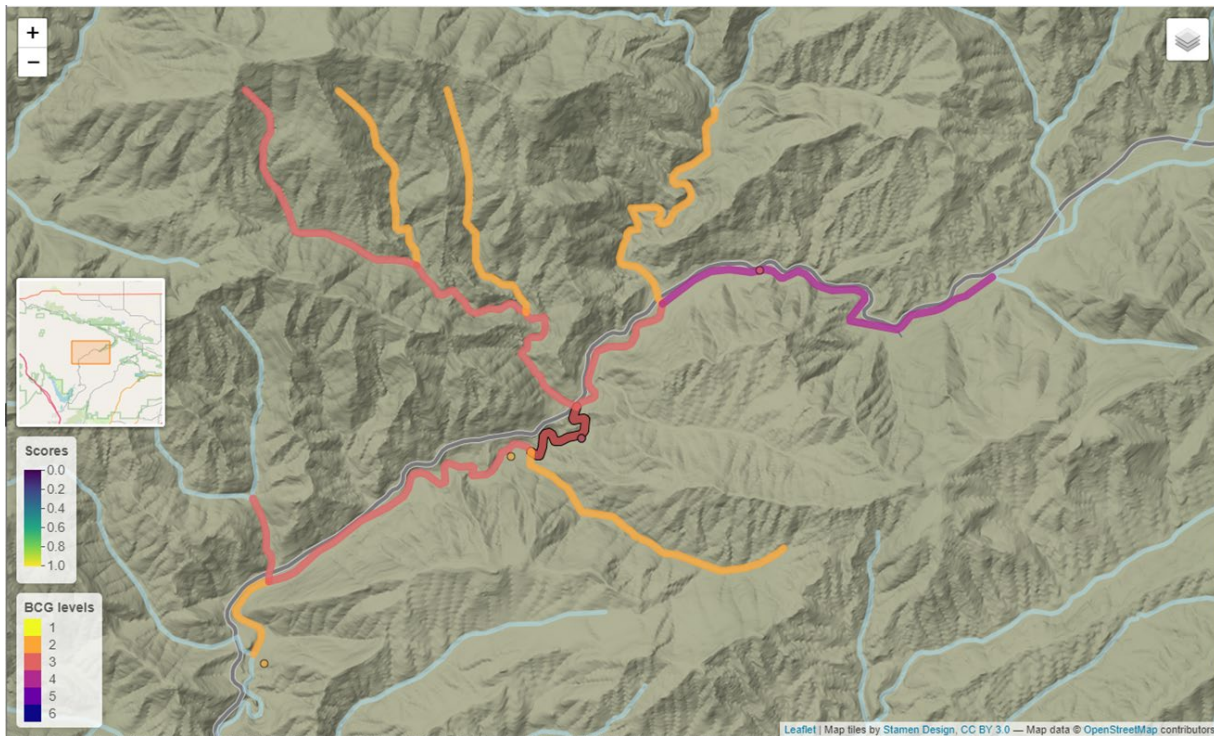
Figure 8. Example of a target reach (dark red color) showing connections with up and downstream reaches and their associated biological condition (CSCI scores) based on observed data or predicted based on the state-wide model. The dots represent BCG levels at specific sites. Understanding the landscape context of a site relative to local biological conditions can help determine a site's priority for restoration or protection.

Threats Subindex

The Threats subindex is made up of two threat types—Future Land Use Change and Fire Hazard (Figure 6). Detailed planned land use geographic data layers for 2050 were obtained from the San Diego Association of Governments SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal. These data layers cover the San Diego region, which were compared to current land uses obtained from the same data source. Projected increases in developed land uses in the reach catchment represent a potential negative effect on future biological condition.

In addition, California's Department of Forestry and Fire Protection, Fire and Resource Assessment Program (CalFire) has modeled and mapped areas of moderate, high, and very high fire hazard based on vegetation, topography, weather, crown fire potential, ember production and movement, and likelihood. CalFire provides statewide geographic data identifying fire hazard (CalFire, 2007). RPPTool incorporates these data for San Diego, Kern, Riverside, Orange, Los Angeles, Ventura, and San Bernardino Counties, where they overlap with the SMC region stream reaches. Fire hazard combines the percent of very high, high, and (optionally) moderate fire hazard area in the catchment, which presents a threat to the success of restoration or protection management actions.

Opportunities Subindex

The Opportunities subindex is made up of three types of information: species conservation, recreational and related co-benefits, and user-defined opportunities (Figure 6). Data derived from the Multiple Species

Conservation Program (MSCP), provides species conservation information for this subindex. The MSCP is an integral part of regional efforts in San Diego County to protect open space and native species. The goal of the MSCP (a 50-year program) is to maintain and enhance biological diversity in the region and maintain viable populations of endangered, threatened, and key sensitive species and their habitats. The MSCP is a cooperative effort among the County, the City of San Diego, and other local jurisdictions, along with the U.S. Fish and Wildlife Service and the California Department of Fish and Wildlife. These public partners work with various private landowners, conservation groups, community planning groups, developers and other stakeholders to carry out MSCP objectives.

The MSCP includes the Multi-Habitat Planning Area (MHPA, which is the County's planned habitat preserve) and Environmentally Sensitive Areas (ESAs). ESAs include conservation and mitigation lands, but also waters that contain rare, threatened, or endangered species, and Areas of Significant Biological Concern. The geographic data are available from the SanGIS/SANDAG GIS Data Portal. Data from each subregion were overlaid on the SMC region catchments to determine the area of protected or preserved land within each catchment. These areas provide an opportunity for co-benefits if a reach within MCSP Plan boundaries is prioritized for restoration or protection.

The Native Aquatic Species Vulnerability Index (NASVI) is also used to characterize the species conservation aspect of the Opportunities subindex. NAVSI is a normalized average of two metrics, the Rarity index (which assesses the average rarity of the species in a watershed, with small-range species being considered rare) and the count of threatened and endangered species in the watershed. Areas with a higher NASVI provide a greater opportunity for co-benefits if the reach is prioritized for restoration or protection.

To characterize recreational and related co-benefits for this subindex, RPPTool incorporates the SanGIS/SANDAG geographic data for current land use, which identifies multiple categories of parks and low-impact recreation areas where recreational co-benefits are most likely to occur. These categories include relatively natural areas such as landscaped greenbelt areas, open space, campgrounds, wildlife and nature preserves, and natural areas with development constraints. They also include more carefully managed and sometimes constructed areas, such as neighborhood parks and recreation centers. The more that a target reach is surrounded by or close to these different types of potential recreational opportunities, the higher the score for the recreational co-benefits indicator. Each potential recreational land use category is calculated as a fraction of the total catchment area for a reach and the contributing categories within each catchment are summed for a total fraction of the catchment that contributes to the recreational co-benefit indicator.

It is understood that a user may have unique knowledge regarding a particular reach with respect to additional benefits or synergies associated with existing or future water resource plans (e.g., Water Quality Improvement Plans [WQIPs]), ongoing studies, and other efforts. User knowledge about sites is often an important aspect of prioritization. To accommodate this, the Opportunities subindex includes a user-provided score to include additional opportunities. The user can add one additional indicator with a score of one (1) and user weights ranging from one (1) to three (3), corresponding with the categorical assignments of low, medium, or high priority, to capture co-benefits of interest.

RPPTool User Input

In the web-based application, the user can select a site from a map populated with all bioassessment sites currently available to obtain the reach identifier or input the identification code for the reach (Figure 9).

Stressor categories and specific stressors, as well as indicator and subindex weights can be altered by the user in the interface to allow flexibility in user prioritization objectives (Figure 10).
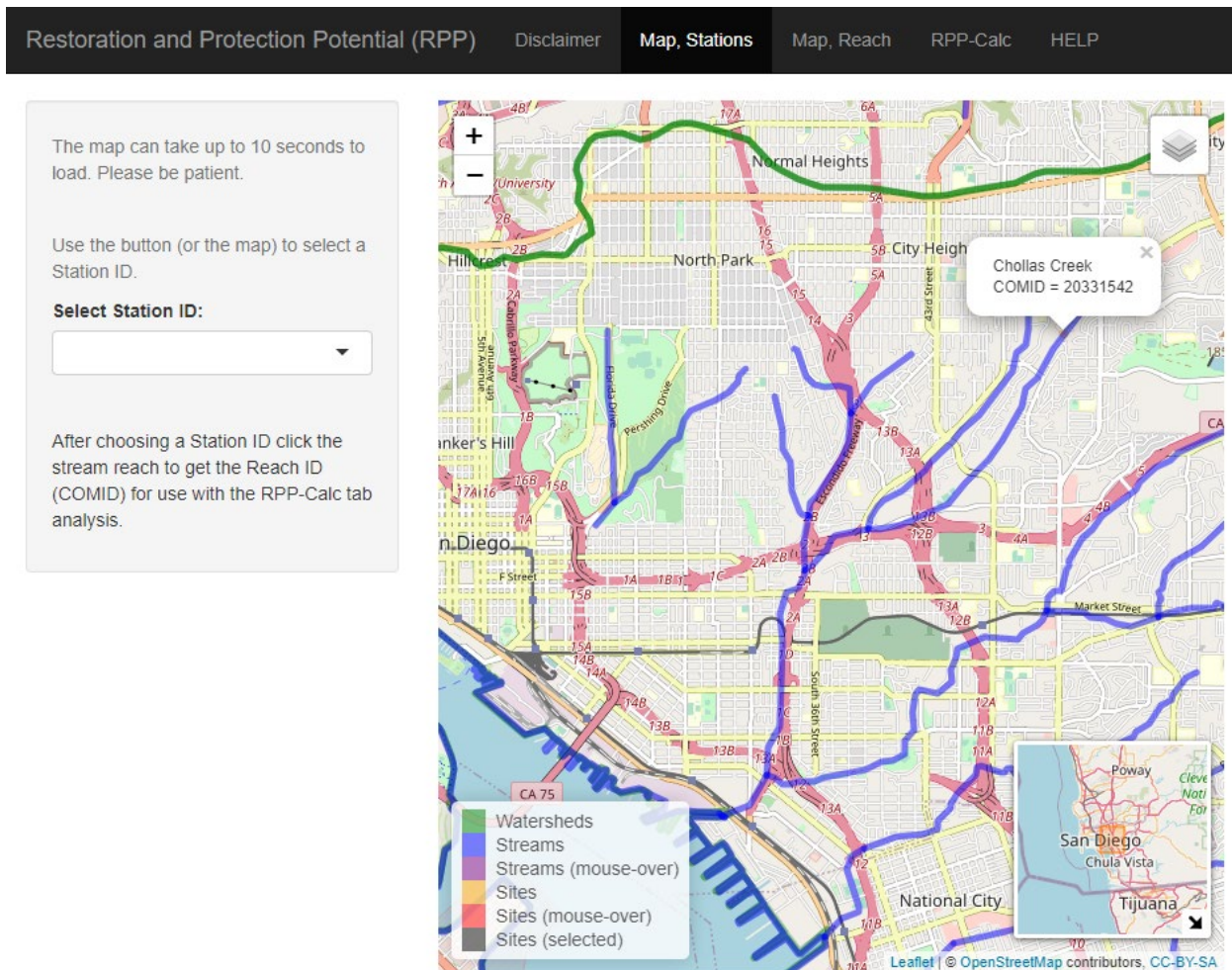

Figure 9. Shiny web application user interface for locating sites

The City *of*
SAN DIEGO


Figure 10. User input page in the RPPTool web application. The user can weight different stressors and other factors in the calculation of the RPP Index for restoration or protection based on their specific objectives.

RPPTool Output

The final output includes a map of the site being analyzed, general information about the site, and whether RPPTool is analyzing the site for restoration or protection (Figure 10). An output table with all results (in the web-application interface or tab-delimited text file) is also presented (Figure 10) that can be downloaded by the user and opened in Microsoft Excel. Subindices are calculated using the available data for each reach in the region and RPP index scores are generated, both for restoration and protection (Figure 11).
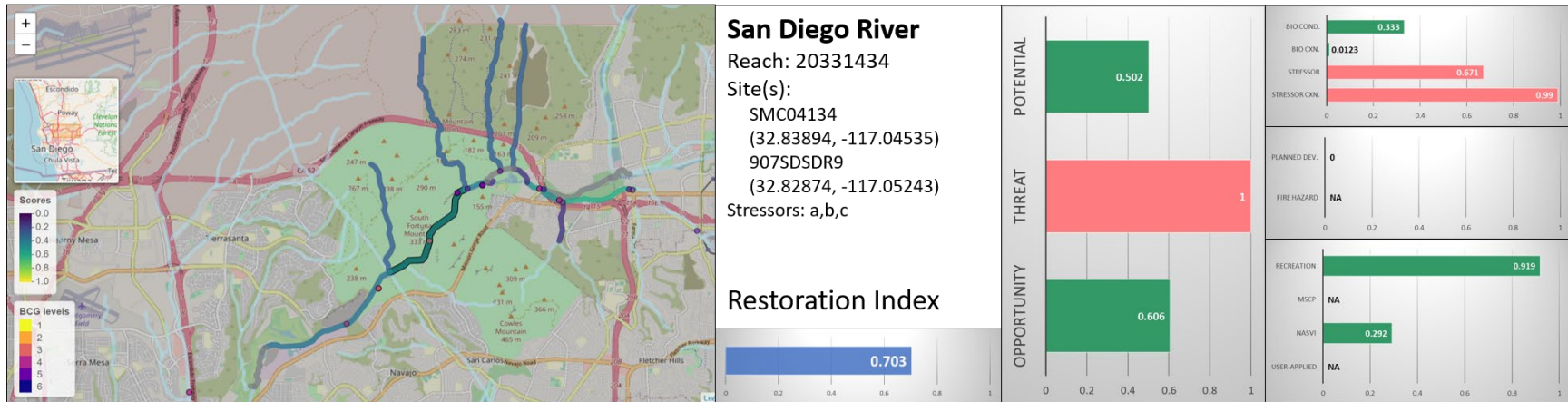
SiteID	COMID	CSCI	Predicted or Observed	Index Type	RPP Index	Potential Subindex	Biological Condition Indicator	Biological Connect.	Stressor	Stressor Connect.	Threat Subindex	Planned Develop	Fire Hazard	Opportunity Subindex	Recreation	MSCP	NASVI	User-applied
SMC04134 907SDSDR9	20331434	0.78	Observed	Restoration	0.703	0.502	0.333	0.0123	0.671	0.99	1	0	NA	0.606	0.919	NA	0.292	NA
907S02774	20333052	0.69	Observed	Restoration	0.592	0.52	0	0.173	0.906	1	1	0	NA	0.255	0.0728	NA	0.437	NA
NA	20333068	0.57	Predicted	Restoration	0.419	0.333	0.333	NA	NA	NA	0.431	0.147	0.991	0.494	0.279	0.263	0.436	1

Figure 11. Example of RPPTool output report in the Shiny web application (*Note: Example results are shown)

4.0 Conclusions and Recommendations

This bioassessment toolkit incorporates the latest scientific advancements regarding California-specific bioassessment data analysis and interpretation approaches, along with recovery potential concepts and other key factors to support the City, Water Board, and other stakeholders within southern California to identify stressors and potential management actions. The analyses and tools incorporated within this toolkit can help provide an indication of those sites or reaches that may be underperforming and could benefit from restoration, as well as those that are in good condition currently but vulnerable to future pressures and needing protection. An important feature of the toolkit is that it considers the landscape context of a reach to determine potential for ecological uplift, potential for threats from planned future urbanization or from fire hazard, and potential co-benefit opportunities that may be realized with restoration or protection.

The combination of CASTool and RPPTool incorporates several other factors that are of interest to stakeholders for prioritizing sites or reaches for restoration and protection including potential habitat for rare and endangered species, as well as species diversity and other conservation-related benefits. In addition, by incorporating options for users to weight a reach more or less in terms of priority based on local knowledge of the reach or surrounding watershed, RPPTool provides opportunities to consider other factors such as socio-economic incentives and synergies with WQIPs, restoration and mitigation efforts (e.g., via the City's Municipal Waterways Maintenance Plan [MWMP]), and other resource management opportunities. The results provide powerful supplemental information to consider along with existing plans (e.g., Watershed Asset Management Plan [WAMP], Watershed Management Plans [WMP], and WQIPs) and other factors in decision making.

Moving forward, the City intends to explore the use of these tools in its planning processes, adding an additional layer of information that as described above, focuses on biological conditions and benefits. The tools will join existing planning tools such as the WAMP, WQIPs, WMPs and the MWMP to help managers determine the best sites for restoration and protection within the City. These tools can also be used by other agencies and practitioners throughout the southern California region. In addition, these tools can help identify priority sites for sampling.

Another advancement is the RPPTool can be used for reaches that currently do not have observational data by using the predicted CSCI distribution based on the state-wide model. That information, along with the corresponding tier on the BCG (also based on the state-wide model), allows the user to examine the potential for ecological uplift within any reach in the SMC region, even if it has not been sampled for a biological assessment.

As noted in this summary report and the Technical Report appendices, CASTool and RPPTool represent a powerful set of screening tools that were designed to meet the objectives of the SEP and other critical needs. While these tools have been shown to provide very useful information to identify data gaps and support management decisions, the results should be carefully considered. Also, currently, data from the SMC portal cannot be accessed automatically in real time via the web. This represents a future enhancement that would benefit water resource managers. It is recommended that the toolkit be updated so that users can take advantage of the most current data and site information available in the region.

While this toolkit currently provides useful information for the City and other stakeholders regarding planning and implementation of stream restoration and protection efforts, additional enhancements are

anticipated through future interactions with the SMC and its member agencies, the Water Board, other stakeholders in the region, and potential development for state-wide use. Examples of enhancements that the Project Team identified which could make the toolkit even more useful include: (1) the ability to run multiple sites or reaches at a time through the RPPTool web application so the user can more readily compare scores among different sites; (2) provide integration with the SMC data portal to automatically update the site information in the toolkit; (3) incorporate additional species-specific stressor responses; and (4) incorporate additional information on perceived threats or opportunities that may become available.

5.0 References

Beck, M., R.D. Mazor, S. Johnson, K. Wisenbaker, J. Westfall, P.R. Ode, R. Hill, C. Loflen, M. Sutula, E.D. Stein. 2019. Prioritizing Management Goals for Stream Biological Integrity Within the Developed Landscape Context. Freshwater Science 38.

CalFire. 2007. Fire Hazard Severity Zones Maps. Sacramento, CA: State of California, Office of the State Marshal. Retrieved from https://osfm.fire.ca.gov/divisions/wildfire-planning-engineering/wildland-hazards-building-codes/fire-hazard-severity-zones-maps/. Accessed 17 March 2020.

California Department of Fish and Wildlife. (2020). *NCCP Plan Summary - San Diego Multiple Species Conservation Program (MSCP)*. Sacramento, CA: California Department of Fish and Wildlife, Habitat Conservation Planning Branch. Retrieved from https://wildlife.ca.gov/Conservation/Planning/NCCP/Plans/San-Diego-MSCP.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2019). Shiny: Web Application Framework for R.R. Package Version 1.3.2. Retrieved from https://CRAN.R-project.org/package=shiny.

City of San Diego, Planning and Development Services. (1997). Multiple Species Conservation Program (MSCP). Retrieved from https://www.sandiegocounty.gov/pds/mscp/. Accessed 29 March 2020.

Mazor, R.D., A.C. Rehn, P.R. Ode, M. Engeln, K.C. Schiff, E.D. Stein, D.J. Gillett, D.B. Herbst, and C.P. Hawkins. 2016. Bioassessment in complex environments: Designing an index for consistent meaning in different settings. Freshwater Science 35(1): 249-271.

Paul, et. al. 2020. Characterizing benthic macroinvertebrate and algal biological condition gradient models for California wadeable Streams, USA. Ecological Indicators 117

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SANDAG/SanGIS. (2015). *SanGIS/SANDAG GIS Data Warehouse*. San Diego Association of Governments (SANDAG) and San Diego Geographic Information Source (SanGIS). Retrieved from https://rdw.sandag.org

SANDAG/SanGIS. (2018a). SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal - Existing Landuse. Retrieved from https://sdgis-sandag.opendata.arcgis.com/datasets/existing-landuse-1. Accessed 21 March 2020.

SANDAG/SanGIS. (2018b). SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal - Planned Landuse. Retrieved from https://sdgis-sandag.opendata.arcgis.com/datasets/planned-landuse-1. Accessed 21 March 2020.

Stein, E.D., R.D. Mazor, A. Sengupta, K. McCune, B. Bledsoe, S. Adams, S. Eberhart, M. Pyne, P. Ode and A. Rehn. 2017. Development of Recommended Flow Targets to Support Biological Integrity Based on Regional Flow-ecology Relationships for Benthic Macroinvertebrates in Southern California Streams. Technical Report 974. Southern California Coastal Water Research Project. Costa Mesa, CA.

City of San Diego. (1998). *Multiple Species Conservation Program (MSCP Plan)*. San Diego, CA: The City of San Diego. Retrieved from https://www.sandiegocounty.gov/content/dam/sdc/pds/mscp/docs/SCMSCP/FinalMSCPProgramPlan.pdf.

USEPA. (2016). *Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems* (EPA/822F-16/002). Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://www.epa.gov/wqc/practitioners-guide-biological-condition-gradient-framework-describe-incremental-change-aquatic.

USEPA. (2017). *Causal Analysis/Diagnosis Decision Information System (CADDIS)*. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development. Retrieved from https://www.epa.gov/caddis.

USEPA. (2018). *EnviroAtlas - Native Freshwater Aquatic Biodiversity by 12-digit HUC for the Conterminous United States*. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory. Retrieved from https://www.epa.gov/enviroatlas/forms/enviroatlas-data-download. Accessed 16 April 2020.

USEPA. (2019). *EnviroAtlas: Native Aquatic Species Vulnerability Index (NASVI)*. Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://enviroatlas.epa.gov/enviroatlas/DataFactSheets/pdf/ESN/Nativeaquaticspeciesvulnerabilityindex.pdf.

USGS and USEPA. (2012). National Hydrography Dataset Plus (NHDPlus Version 2. Medium Resolution for California). Retrieved from https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus. Accessed 11 November 2012.

Appendices

Appendix A. CASTool Technical Memo

The City of SAN DIEGO

MEMO

To: Vicki Kalkirtz, Ruth Kolb (City of San Diego)

From: Clint Boschen, Jerry Diamond, Ann Roseberry Lincoln, Erik Leppo (Tetra Tech)

Subject: Causal Assessment Screening Tool (CASTool) for the Supplemental Environmental Project (SEP). Restoring and Protecting Our Waterways: Bio-assessment Tools and Priorities for Action

Table of Contents

Figures

Tables

1.0 Introduction

As part of the process for meeting future requirements of the Regional Board's proposed biological objectives Basin Plan Amendment, the City of San Diego has been developing a screening causal assessment framework that can rapidly screen stream sites where biological assessment data are available. The Regional Board has included screening causal assessment in their proposed Basin Plan Amendment to help permittees and others address restoration opportunities in biologically impaired streams. A rapid screening tool would help analysts both (1) identify likely causes of biological impairment in a given waterbody and (2) rule out pollutants (e.g., certain pesticides) or other types of pollution (e.g., habitat constraints) that are unlikely to cause impairment, thereby focusing restoration and protection actions. The Project Team, which includes City staff, Tetra Tech, SCCWRP, and staff from the Regional Board, developed a causal assessment screening tool (CASTool) in R (R Core Team 2019) that can accomplish these tasks for non-ephemeral waterbodies in the Stormwater Monitoring Coalition (SMC) bioassessment region. This report outlines the screening causal assessment (SCA) framework and how it is executed by the CASTool. The CASTool provides results that indicate which stressor(s) are likely causes of biological impairment. This information, along with assessments regarding recovery potential and additional co-benefits, can be used to help identify restoration and protection priorities and inform management actions to reduce the stressor(s) and achieve the desired ecological lift.

The CASTool approach focuses on the major lines of evidence that are routinely conducted using available site data but has the advantage of automating site analyses and presenting the weight of evidence for or against each likely candidate stressor to identify likely causes of biological impairment. The CASTool uses an adaptation of the comparator site approach conceived in USEPA's Causal Analysis/Diagnosis Decision Information System (CADDIS, https://www.epa.gov/caddis [USEPA 2017]).

The causal assessment screening approach needs to be capable of identifying the likelihood of different potential stressors as causes of observed biological impairment at a site. To accomplish this, the CASTool includes a framework to streamline and automate selected analyses within EPA's CADDIS guidelines. The automated analyses incorporate altered flow metrics; stressor-specific diagnostic indicators, including benthic macroinvertebrate tolerance values for specific water quality stressors (e.g., fine sediments and conductivity); and a transparent scoring system for summarizing which stressors are the most likely causes of biological impairment.

2.0 The Screening Causal Assessment Framework

2.1 Comparator Site Approach

Comparator sites are used to assess conditions at target (impaired) sites and to help identify the potential causes of impairment. Traditionally, the causal assessment process considered relatively simplistic environmental scenarios, in which there is a single stressor (e.g., an industrial facility discharge) and analogous sites upstream of the stressor source that are otherwise relatively unaffected by anthropogenic activities (e.g., reference sites). However, in the Stormwater Monitoring Coalition (SMC) region of southern California (and many other regions of the U.S.) it is common to have multiple watershed scale non-point and point sources of potential stressors that affect biological conditions at a site depending on the extent of human activity and alterations to the landscape. As a result, it is challenging to easily identify appropriate individual sites to use for comparison.

The SCA framework refines the comparator site methodology based on harmonizing two existing methods: (1) physical setting approach using cluster analysis based on a set of abiotic factors most closely related to CSCI scores and components metrics in the SMC region, and (2) grouping sites based on taxa dissimilarity using the State-wide Observed/Expected (O/E) model. Previous to the SEP, the City's comparator site framework used a statistical approach to identify different groups of comparator sites (clusters of sites; Figure 2-1) based on non-biological (i.e., abiotic) factors that are known to affect distribution of benthic macroinvertebrate species and other aquatic life, such as elevation, slope, and underlying geology.

To determine which comparator site approach should be incorporated in the SCA framework the Project Team tested many biological sites in the SMC region using four evaluation criteria:

1. There are enough sites to support causal assessment analysis
2. The group of comparator sites represent a gradient in biological condition in relation to the test site; specifically, at least some comparator sites are in better condition than the test site.
3. The comparator sites potentially support a benthic macroinvertebrate assemblage comparable to that of the test site in the absence of human or natural disturbance
4. The range of natural gradients within the comparator site group include values observed at the test site.

Results of these analyses are detailed in Appendix A: Year 1 Technical Memo appended. The physical setting approach performed almost identically with O/E approach, however, it was acknowldeged that both approaches have advantages and limitations. Therefore a hybrid comparator site approach was selected for the CASTool because it incorporates both modeled expected taxa and abiotic factors important in the SMC region. The CASTool utilizes groups of comparator sites based on abiotic factors in a cluster analysis which are then filtered using the O/E model. The cluster analysis helps account for natural variation that might otherwise obscure relationships between biological condition and sources of impairment.

2.2 Inclusion of Flow Ecology and Biological Diagnostic Information

Flow ecology stressor information, based on hydrologic alteration metrics developed by SCCWRP, is incorporated into the CASTool. Table 2-1 summarizes the flow ecology metrics provided by SCCWRP for benthic macroinvterbrates and benthic algae. Metric data for biological sites in the SMC region were provided by SCCWRP for macroinvertebrates and algae separately using the State-wide hydrologic model adapted for the SMC region. Some biological sites did not have flow metric data available at this time due to insufficient information with which to run the hydrologic model. However, many biological sites had modeled flow data, which are incorporated in the CASTool for algae and macroinvertebrates.

Table 2-1. Priority hydrologic metrics and associated thresholds used in the regional flow-ecology relationships. Metrics are grouped by the hydrograph component they represent. Thresholds are expressed as the change in metric value (delta H) associated with poor biological condition (CSCI <0.79). Metric effects on biology were typically strongest during either average, wet, or dry rainfall years, or all years combined (overall). NT = no threshold established (from Stein et al 2017)

Hydrograph component	Metric	Metric definition	Critical precipitation condition	Decreasing threshold	Increasing threshold
Duration	NoDisturb (days)	Median annual longest number of consecutive days that flow is between the low and high flow threshold	Average	-64	NT
	HighDur (days/event)	Median annual longest number of consecutive days that flow was greater than the high flow threshold	Wet	−3	24
Magnitude	MaxMonthQ (m^3/s)	Maximum mean monthly streamflow	Wet	NT	1.5
	Q99 (m^3/s)	99th percentile of daily streamflow	Wet	−0.01	32
Variability	RBI (unitless)	Richards-Baker index of stream flashiness	Dry	NT	0.25
	QmaxIDR (m^3/s)	Interdecile range of flow	Overall	-5	2.5
Frequency	HighNum (events/year)	Median annual number of events that flow was greater than high flow threshold	Dry	NT	3

In addition to altered flow stressor information, several diagnostic refinements were also explored in this SEP, including tolerance of macroinvertebrates to excess nutrients, altered flow, and fine sediments. Taxa tolerance information was sufficiently robust for fine sediments and conductivity/total dissolved solids to include in the CASTool as biological diagnostic lines of evidence (See Appendix A: Year 1 Techncial Memo).
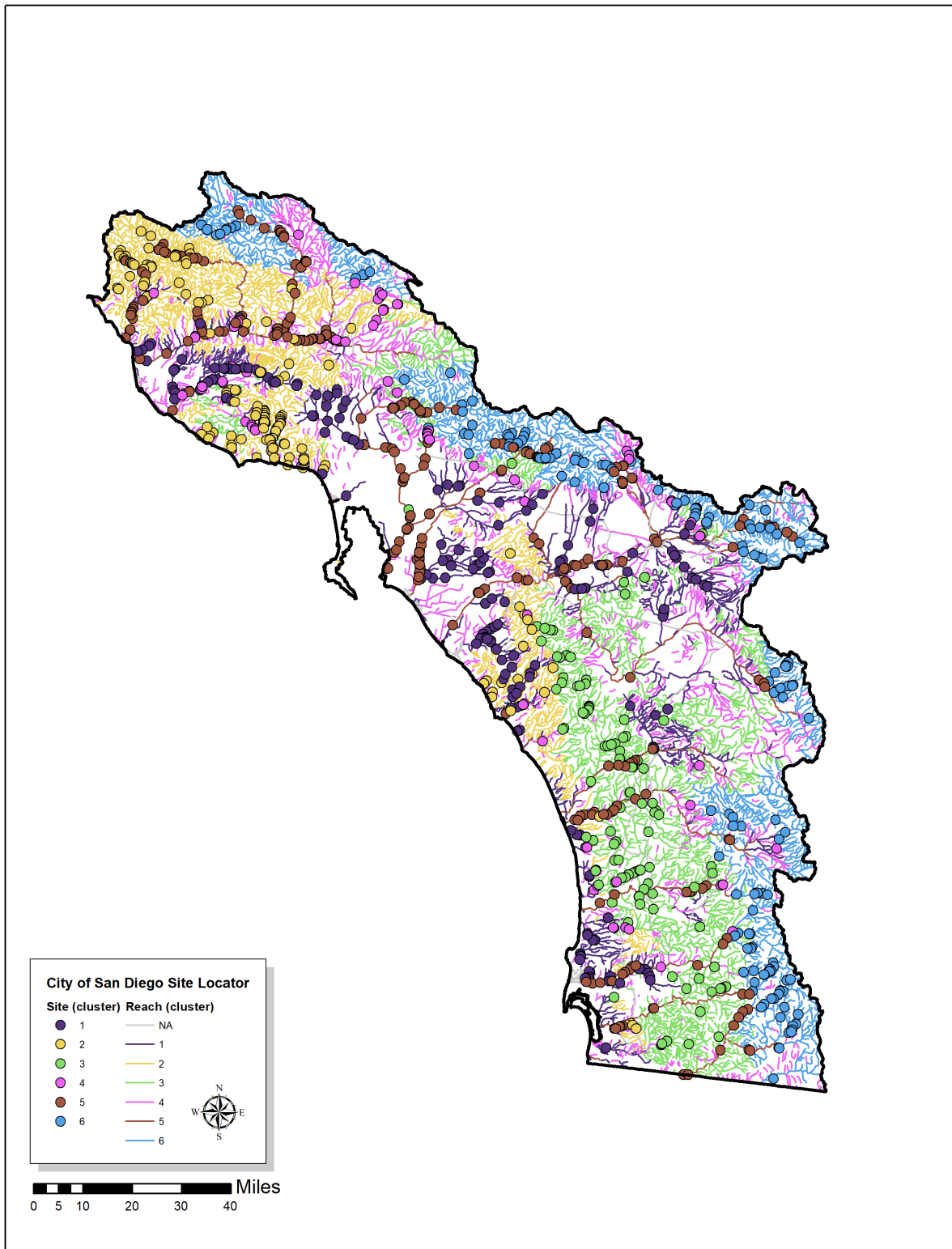
Figure 2-1. Sites from the comparator site analysis for the SMC region in southern California. Site clusters in this figure are based on physical similarity.

3.0 Data Processing Using the CASTool

The CASTool consists of multiple functions, input data files as described in this Chapter, and a "skeleton" code that invokes the functions in the proper order, with little user input required. These functions are written in the Statistical Programming Language R (R Core Team 2019) and contained in a package. Additionally, a user interface was developed in shiny (see Chapter 5). The R package, shiny, facilitates development of R-based web applications, allowing anyone to use the CASTool package even if they have no R experience to operate the code functions. The details of the analytical portions of the tool are described in this section. The font in blue, e.g. `exampleFont`, indicates data objects or functions that are part of the CASTool. Appendix B presents the data dictionary for data files used as input to the CASTool.

3.1 Data

In addition to the reach-based cluster data files prepared during the clustering process, water quality, physical habitat metrics and indices, benthic macroinvertebrate and algal abundance, metrics, and indices (i.e., pMMI and associated metrics; ASCI and associated metrics), and flow metrics most likely to affect benthic macroinvertebrates and algae, are inputs into the CASTool. Site-specific data and metrics, and flow metrics were received from SCCWRP and Regional Board staff. This information is incorporated into the CASTool. Stressor and response data received were reviewed and standardized for use in the CASTool. Stressor names were standardized, and result values converted to common units.

Two separate types of stressor data are available and were collated into two separate sets of files: one for measured data (chemistry and physical habitat) and one for modeled data (flow metrics). For measured data, multiple result columns for stressor data were retained, including the result value as originally reported (e.g., "<2.0"); the result as a numeric value where non-detects are not included; and the result value for which non-detects are converted to half the detection limit, if a detection limit was provided. The current default for the CASTool is to use numeric values without non-detects.

3.1.1 Input Data Files for the CASTool

Table 3-1 summarizes the types of stressor and biological data used in the CASTool. The approach used to code the CASTool provides a great deal of flexibility so that additional stressor and biological parameters could be readily incorporated in a future version of the CASTool if desired.

Table 3-1. Summary of stressor and biological input data currently incorporated in the CASTool.

Stressor Group	Stressors
Habitat	Index of physical habitat integrity (IPI)
	Component metrics of the IPI
Altered flow	Flow ecology metrics for SMC region
Nutrients	e.g., total and dissolved nitrogen, phosphorous
Water Quality	General parameters such as dissolved oxygen, conductivity, pH, alkalinity, dissolved solids, and temperature; ions, toxics (pesticides, metals)
Macroinvertebrates	CSCI

Stressor Group	Stressors
	pMMI
	Component metrics of CSCI
	Taxa counts in the sample
Algae	ASCI
	Component metrics

The following lists the different types of information used and where they reside in the CASTool.

- Comparator site information
    - o SMCClusters.tab – Indexed by COMID (unique identifier of the segment in the NHD+ shapefile) includes cluster assignment for all SMC sites and transformed and imputed StreamCat variables used in the clustering algorithm.
    - o SMCBCDist.tab – A Bray-Curtis dissimilarity distance matrix, with each site identifier in a single column as row names and in a single row as column names. The diagonal contains zero values.
- Site information
    - o SMCSitesFinal.tab – Indexed by site identifier, includes site physiographic information (latitude, longitude, elevation category, COMID in which the site is located), and cluster membership, along with other relevant site-specific information.
    - o SMCSiteSummary.tab – Indexed by site identifier and sample collection date, includes sample identifiers for multiple sample types (water quality/chemistry, physical habitat, benthic macroinvertebrate, and algae).
    - o SMCCoOccurData.tab – Indexed by ChemSampleID, provides information relating the stressor sample with the response sample, including numeric data.
- Stressor information
    - o SMCStressorDataFinal.tab – Indexed by sample identifier and analyte, includes text-based result values (e.g., "<2.0"), numeric result values for which non-detects are blank, and result values for which non-detects are present and calculated to be one-half the detection limit for that analyte in that sample.
    - o SMCStressorInfoFinal.tab – Indexed by analyte, includes metadata about each analyte, such as the sample fraction, standard units, the stressor group to which the analyte belongs, whether or not the analyte should be included in the quantitative screening causal assessment, whether or not the results should be log-transformed, whether or not laboratory toxicity-based species sensitivity distributions are available, and whether or not stressor-specific tolerance values are available and which of those tolerance values indicate sensitivity or tolerance. A final field indicates whether stress increases with increasing or decreasing stressor value.
    - o SMCModelDataFinal.tab – Indexed by sample identifier and parameter, includes numeric values in the same format as the SMCStressorDataFinal.tab, where the text results, numeric results, and numeric results with non-detects calculated to be one-half the detection limit are all equal, because finite numeric values are available for every parameter.

- o SMCModelInfoFinal.tab – Indexed by parameter, this file follows a parallel structure to SMCStressorInfoFinal.tab, with multiple fields containing null values. This is the required metadata for the flow metrics.
- Biological Response information
  - o Benthic macroinvertebrates
    - ▪ SMCBenthicCountsFinal.tab – Indexed by benthic sample ID, provides the final taxonomic identification, total number of individuals counted, total number of individuals after adjusting for subsampling, and relative abundance of the final taxon relative to the total sample.
    - ▪ SMCBenthicMasterTaxa.tab – Indexed by final taxonomic identification, includes metadata relevant to the taxon, including phylogenetic hierarchy, functional feeding group, habit, tolerance to anthropogenic stressors, and stressor-specific tolerance values, as available.
    - ▪ SMCBenthicMetricsFinal.tab – Indexed by benthic sample ID, provides the final CSCI score, as well as the O/E and pMMI scores, and all the metrics calculated to obtain the pMMI.
  - o Algae
    - ▪ SMCAlgaeCountsFinal.tab –Indexed by algal sample ID, and will provide the sampling method, final taxonomic identification, lifestage distinct flag, total number of individuals, and relative abundance of the taxon relative to the total sample.
    - ▪ SMCAlgalMasterTaxa.tab –Indexed by final taxonomic identification, and will include metadata relevant to the taxon, including phylogenetic hierarchy and membership in a variety of indicator classes (e.g., pollution tolerance, pH, nitrogen uptake, saprobity, etc.).
    - ▪ SMCAlgaeMetrics.tab –Indexed by algal sample ID, and will provide the value for each sample's Algal Stream Condition Index (ASCI) and the metrics that comprise the ASCI.
- GIS Data
  - o Reaches
    - ▪ SMCReaches.shp – A flowline shapefile, clipped from the NHD Plus flowline file and corresponding with stream reaches within the Stormwater Monitoring Coalition (SMC) region.
  - o SMC region outline
    - ▪ SMCBoundary.shp – A polygon shapefile outline of the SMC region.
- Optional data
  - o SMCImpairments.tab – Data specifying known impairments, as described in the California State Clean Water Act Section 303(d) list of impaired waters. The data include the COMID, which matches the COMID on the site list for listed waterbodies only, the waterbody name, the pollutants listed, and the final listing decision.
  - o SMCMods.tab – Indexed by COMID, this file indicates the reach modification status and type of modification resulting from the GIS-based screening model.

3.2 Lines of Evidence Analyzed

The CASTool incorporates analyses of several lines of evidence that have been identified in US EPA's CADDIS and that often have data available from bioassessments to enable analyses (Table 3-2). Not all lines of evidence that are in the CADDIS framework can be evaluated using an automated screening causal assessment tool such as the CASTool. This is mostly due to a lack of data and other information with which to analyze certain lines of evidence in CADDIS (e.g., manipulations involving experimentation with current conditions). The lines of evidence listed in Table 3-2 are most frequently used in traditional causal assessments and provide the strongest evidence based on the types of stressor and biological data routinely collected. Future versions of the CASTool could incorporate other lines of evidence if desired and data become available.

Table 3-2. Lines of evidence identified in U.S. EPA's CADDIS framework that are incorporated in the CASTool.

Line of Evidence (from EPA CADDIS)	Question addressed
Spatial/temporal co-occurrence	Does the biological effect occur in the same location and at the same time as the stressor and not when or where the stressor is absent?
Verified predictions	Given the stressor's mode of action, is a predicted response observed?
Stressor-response relationships from the field	Does the adverse response increase or decrease as the stressor increases or decreases?
Temporal sequence	Does the cause precede the adverse effect?
Stressor-response relationships from other field studies.	Using data from outside the case, does the adverse response increase or decrease as the stressor increases or decreases, and are the candidate cause levels at the target site sufficient to cause similar impairment?

The CASTool comprises a series of functions written in the R Statistical Programming Language (v4.0.2, requires v3.5.1 or greater) to process and analyze data and lines of evidence. An overview of the CASTool is shown pictorially in Figure 3-1. The CASTool package functions generally utilize either base R methods or tidyverse methods. Graphs typically are created using ggplot2. Tetra Tech has separately written a skeleton program that iterates over a series of target sites provided as an Excel file and calls the methods of the CASTool package in the appropriate order to perform a full screening causal assessment in batch mode. Additionally, Tetra Tech has designed an interface in shiny, an R-based application for interactive data analytics, allowing the user to perform a full screening causal assessment on an individual site.

The first functions are descriptive, while the remaining functions are designed to identify potential stressors at the target site and to evaluate the causal lines of evidence for them. The descriptive functions include: getSiteInfo, getDataSets, getClusterInfo, getComparators, getOutliers, getQualSites, and getStressorList. The analytical functions include: getBioMatches and getBioStressorResponses, getCoOccur, getTimeSeq, and getVerifiedPredictions. The final three functions are getWoE (weight of evidence), getReport (reporting summary), and getSummaryAllSites (combine weight of evidence from all sites in the results folder). Of these final functions, the first pulls together all the detailed scores generated by each of the analytical functions evaluated for the target site, and the second pulls all the results from each of the descriptive functions and selected analytical output generated for the target site. The last combines the output of getWoE for each site into a complete list of results for all sites. These combined files are used (if desired) in the companion RPPTool. Additional functionality is provided in the skeleton code structure. Each of these functions are described in more detail below.
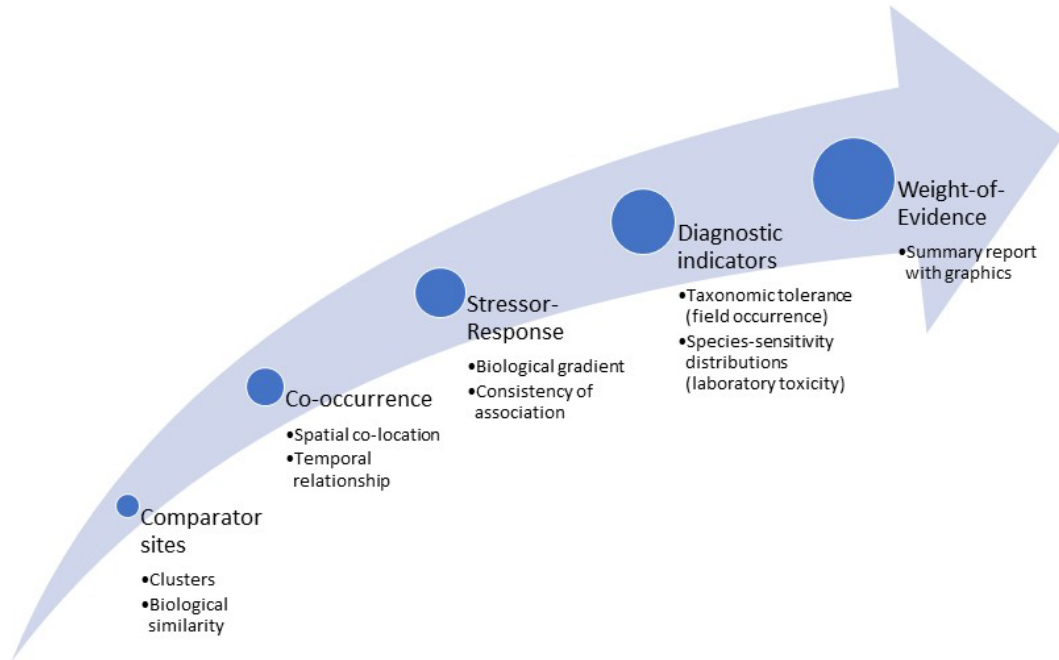
Figure 3-1. Graphic conceptual diagram of the modules in the CASTool. Stressors include: water quality, pollutants, toxics, altered flow, and habitat alterations. Responses include: benthic macroinvertebrate and algae taxa characteristics.

3.3 Prepare the Data

After the input data files are read into the CASTool, some data preparation steps are required. These steps are not taken prior to reading the input data, so that newly obtained data do not require significant preparation other than standardizing the file format and meeting domain requirements (limited data types such as integers, dates, and parameter names that should be standardized). These steps include flagging outliers so that the user can determine whether or not to exclude them, and preparing stressor/response "matched" data within a specified number of days to account for samples not collected on the same day (default is 10 days).

`getOutliers.` This function flags outliers using two methods. If a data point is identified by both methods as an outlier, it is flagged to be removed if the user opts to remove outliers (default is to remove them). The first method flags all data points falling 1.5 times outside the interquartile range on either the high or low side. The second method flags all data points falling outside 6 standard deviations from the mean. Using both of these methods and requiring both to be true provides a conservative method for eliminating outliers. The data points meeting these criteria will likely be extreme values representative of data entry or collection errors.

`getCoOccurDataset.` This function compares stressor samples obtained from bioassessment sites with each type of response data (benthic macroinvertebrates and algae separately) to pair samples meeting the criteria of the stressor sample being obtained prior to the response sample, but not greater than a specified number of days earlier. This concept of "lag days" accounts for the fact that while most site assessment protocols require completion on the same day, that is not always possible due to time or weather constraints in the field. The default number of lag days is 10. In other words, a stressor and response sample are considered paired if they were obtained not more than 10 days apart, with the stressor sample obtained prior to the response sample at the same location. It is important to note that modeled flow metrics are not obtained on a specific date, and all modeled flow metrics are assumed to be paired with all response data from the same location.

3.4 Define the Case

A target site is a specific bioassessment location with a biological impairment, for which the CASTool user seeks to identify likely causes of impairment, or an unimpaired site, for which the CASTool user seeks to identify potential stressors to which the site may be vulnerable. "The case" is defined as the target site and its comparators.

Cluster analysis identifies a larger group of similar sites that can serve as comparators for the screening causal assessment. Clusters are identified using physical characteristics such as geology, lithology, hydrology, and climate, as discussed in Section 2. Cluster data are generated at the reach scale. A target site inherits the cluster membership from the reach on which it is located. Bioassessment sites that are located on reaches belonging to the same cluster comprise the initial comparator set for a given target site. The

sites identified as belonging to the same cluster are then filtered based on expected taxa dissimilarity using the State-wide CSCI model to identify the final set of comparator sites for a given target site.

The following functions are used to define the case:

getClusterInfo. This function prepares graphical comparisons of cluster characteristics for each of the clusters in the dataset, with target site and reference data overlaid onto the clusters to which they belong. Graphics include box plots showing the distribution of each StreamCat variable in the cluster files (imputed, transformed values) to which the target site belongs. Values for reaches containing reference sites are overlaid in blue, and the value for the reach containing the target site is overlaid in red. Figure 3-5 shows an example.

getComparators: Once the cluster membership has been defined for the test site, the clustered comparator sites are then filtered based on the Bray-Curtis dissimilarity matrix of the expected taxa between each comparator site and the test site using the State-wide O/E model (discussed in Section 3.2.2 below). Only fairly common taxa (taxa that have at least a 50% [0.5] probability of capture at each site) are included in the expected taxa filter of comparator sites. The result is a subset of the cluster sites which are then taken forward in the CASTool as the comparator sites for the test site. This function identifies comparators based a default dissimilarity distance of 0.05 (95% of all common taxa are the same), provided a minimum of 30 sites qualify. If fewer than 30 sites have similar expected biology, additional sites are added until the 30-site minimum is achieved. The function generates a table of comparator sites, along with their dissimilarity distance (0 is 100% similar, 1 is 100% different), sample identifiers obtained from the comparator sites, along with their CSCI score and a statement conveying whether or not the sample is considered degraded, and any flags associated with the sample obtained from the CSCI calculation (Mazor et al. 2015).

getSiteInfo: This function identifies site characteristics for the target site, including stressor and response samples collected from the site, the reach identifier on which the site is located, the cluster to which the reach (and therefore the site) belongs, any 303d-listed impairments if available, and reach modification status if the mods file exists. The function generates an image file used by the web interface containing a map of the study area (e.g., Figure 3-2, top image), the Station ID of the station selected, and the stream flowlines. When run in batch mode, the map image (Figure 3-2, bottom image) shows the SMC region, all stream flowlines, all sites in the region (grey dots), sites in the same cluster as the target site (teal dots), reference sites (dots with blue outlines), and the target site (red triangle). This function also collates any site photos from a centralized location into the site results folder, prepares boxplots of biological indices and metrics compared to comparator sites (Figure 3-3), and bar plots of anthropogenic landscape stressors at catchment and watershed scales (Figure 3-4).

SMC04134
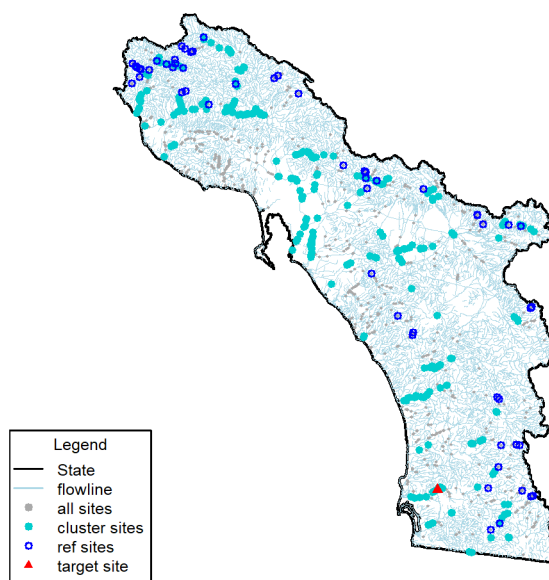
Figure 3-2. Top: Map from CASTool web application showing location of SMC04134 station (grey dot) and flow lines (blue). Bottom: Map from the CASTool run in batch mode, showing the target site (red triangle), SMC boundary (black line), all flowlines in the SMC region (blue lines), all sites (grey dots), sites in the same cluster as the target site (teal), and reference sites (blue outlines).
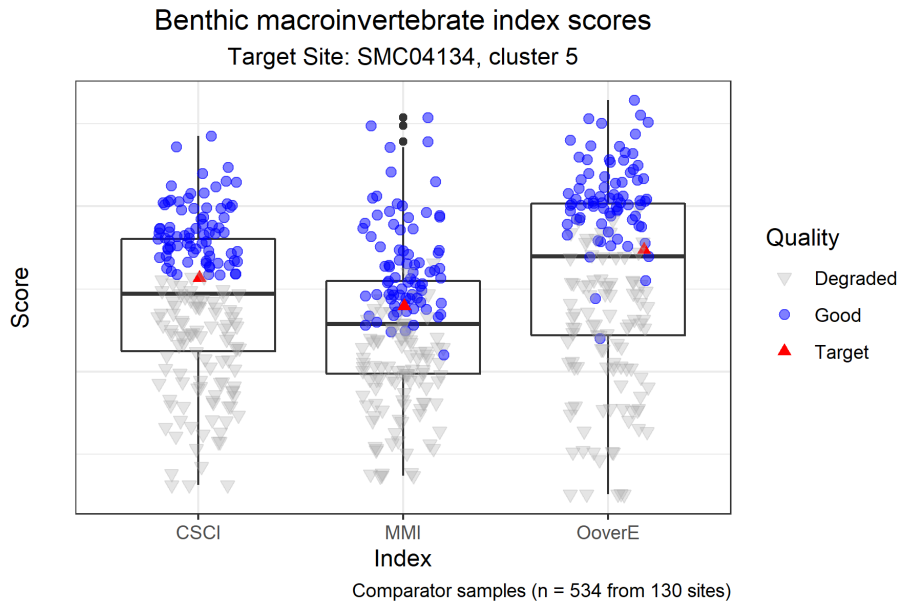
Figure 3-3. Boxplot showing target site biological metrics relative to comparator site metrics for all available samples. Blue dots represent not degraded samples, grey triangles represent degraded samples, and red triangles represent samples from the target site.
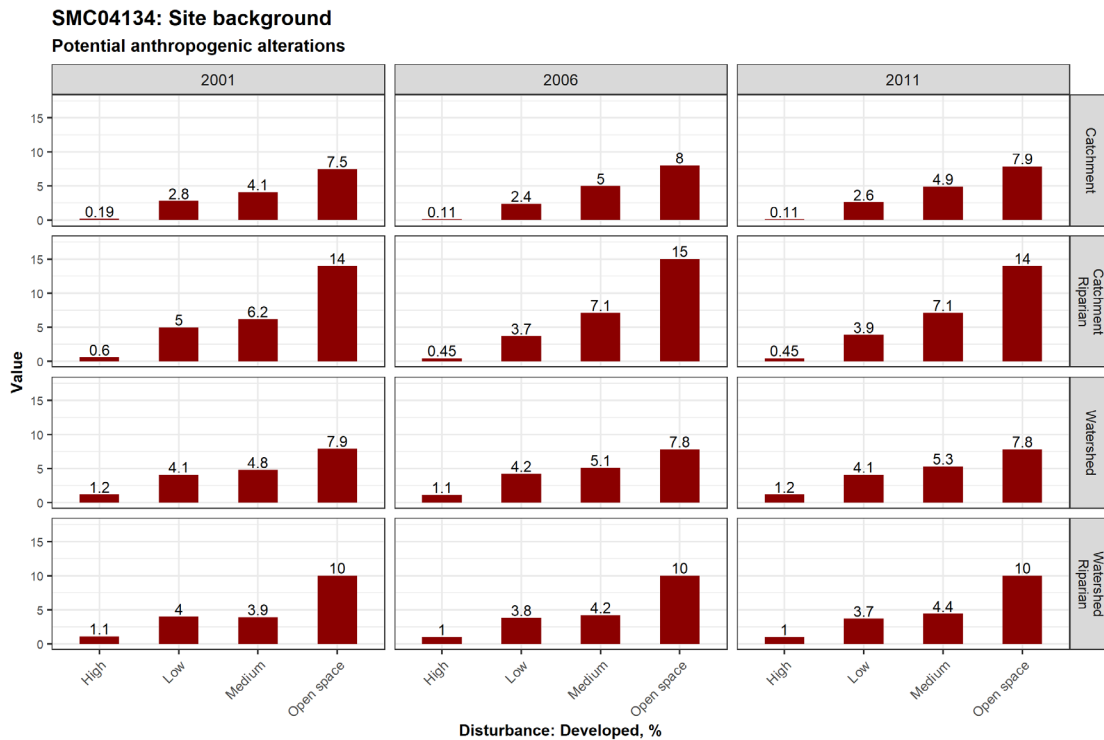
Figure 3-4. Graphic representation of changes in developed land cover in the catchment, catchment riparian zone, watershed, and watershed riparian zone from three NLCD data sets (2001, 2006, and 2011).

3.5 Identify Candidate Causes

getDataSets. This function prepares data subsets based on reported stressor data from the target site to determine which reported stressors represent candidate causes that should be evaluated further. This function first identifies any stressors ever detected at the target site, and subsets the entire stressor data set to include only those detected stressors. A second subset is created, comprising samples from comparator sites for the target site. This last subset is created following the application of the Bray-Curtis dissimilarity distance matrix and therefore contains a final, smaller, set of comparator sites.

getStressorList. This function looks at the distribution of each analyte among sites belonging to the same comparator group as the target site and determines which analytes at the target site are reported at values exceeding a specified threshold, expressed as some percentile of the distribution. For analytes demonstrating greater stress at higher values, an upper percentile is used as the threshold. For analytes demonstrating greater stress at lower values (e.g., pH and dissolved oxygen), a lower percentile is used.

The function writes a series of box plots for the comparator sites. Each box plot graphic corresponds with one stressor group within a stressor type (e.g., turbidity and suspended solids are within the group of the water quality type) similar to the example provided in 3-4. Box plots for each potential stressor are normalized, and the interquartile range (IQR) and median shown within the box. The horizontal lines extending beyond the box represent ±1.5 * IQR and outliers are depicted as black circles. Reference site data for each potential stressor are overlaid in blue and the target site data are overlaid as red triangles. For example, in Figure 3-4, fine sediments in the riffle and the reach are relatively high compared to comparator or reference sites suggesting that fine sediments are a candidate cause of biological impairment.
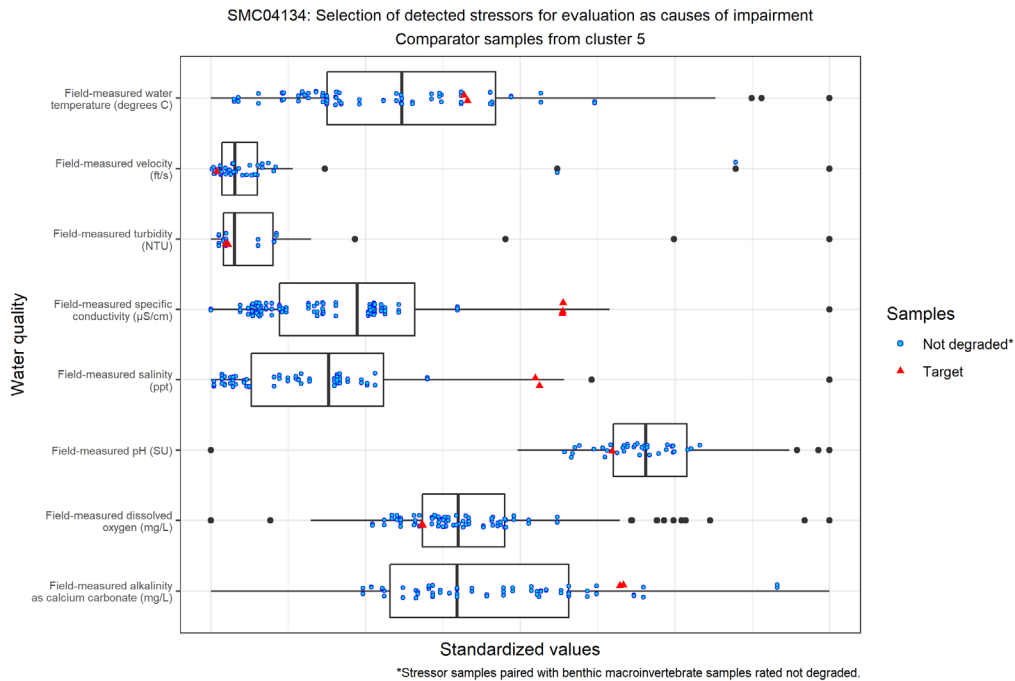
Figure 3-5. Sample box plot produced by `getStressorList`

3.6 Evaluate Data from the Case

`getTimeSeq.` This function plots the value of a stressor over time above the response values over time to discern any temporal patterns, regardless of whether or not the samples are paired. Few sites in the data set have more than one sampling event, however, with the result that frequently this line of evidence cannot be fully evaluated. Figure 3-6 shows an example of a time series plot for one site that had two separate biological samplings and multiple dissolved oxygen measurements within the same time period.

`getCoOccur.` This function evaluates whether greater levels of a potential stressor (i.e., higher levels in the case of chemical pollutants or lower levels in the case of parameters such as dissolved oxygen) occur where and when the biological effect occurs. The biological effect is defined as reduced biological integrity as measure by the CSCI. The first part of this function evaluates the co-occurrence line of evidence. This line of evidence uses the subset of comparator samples with better biology than the target site, determined as
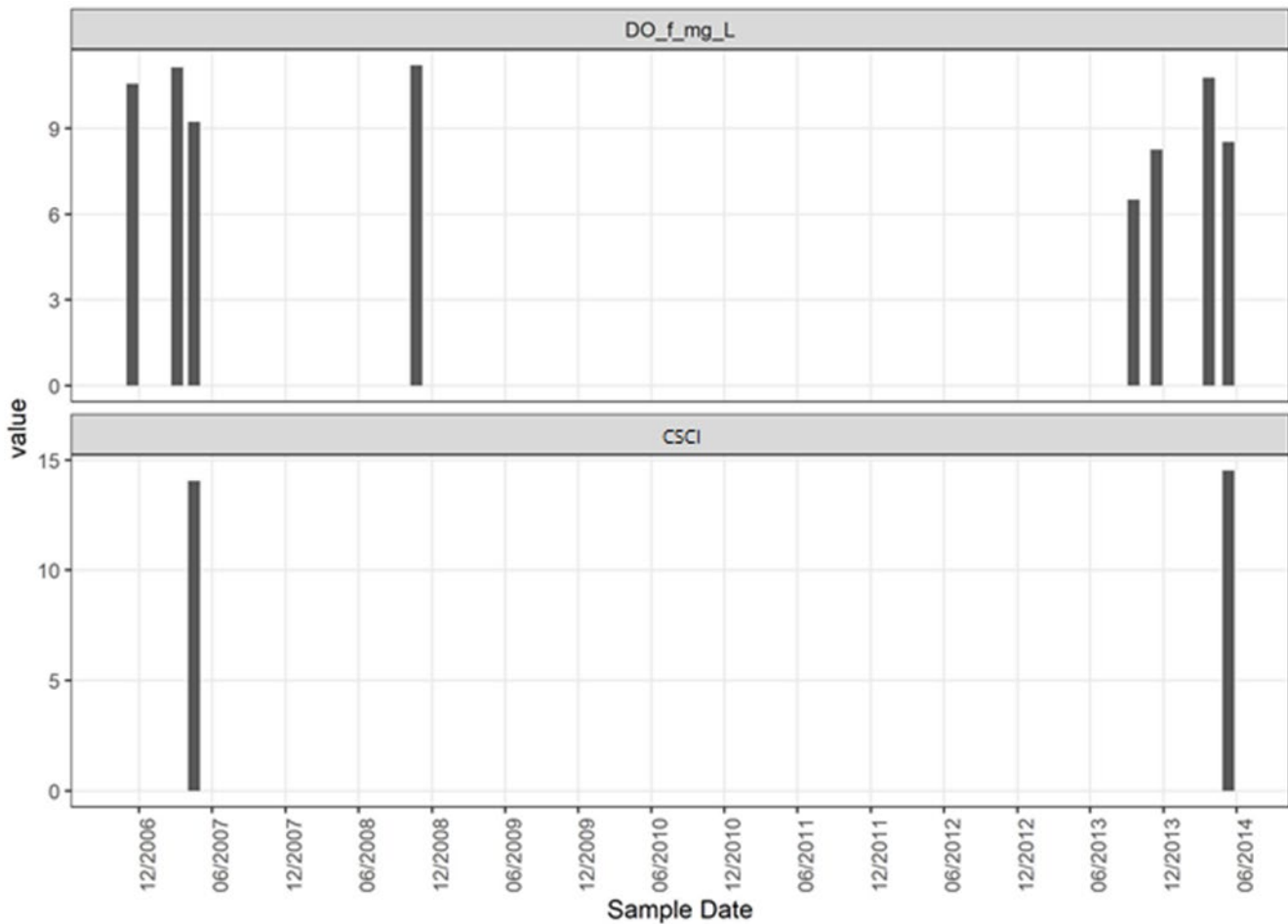
Figure 3-6. Example bar plots generated by `getTimeSeq`.

samples with an index score better than the minimum index score from samples obtained at the target site. The distribution of "better biology" samples is plotted as a box plot, and the target site samples are compared with the distribution (Figure 3-7). The white box shows the interquartile range (IQR, the edges of the box), the median (the black line inside the box), and the horizontal lines to either side represent ±1.5 * IQR. Dots falling outside the horizontal lines represent outliers. The blue dots represent degraded samples whereas grey triangles represent samples that are not degraded. In this example, conductivity recorded at sites with degraded biology (4 blue dots) is shown with conductivity at multiple comparator sites that do not have degraded biology. "Degraded biology" is based on the value determined by the regulating agency as degraded, and that numeric criterion is supplied to the function as an argument. The black dashed lines bracket the "indeterminate" range. The red vertical dashed lines represent the stressor values for each sample obtained from the target site where there is also a paired biological sample. The number of samples represented, and the scores assigned to each target site sample are shown in the bottom right corner of the graphic. For stressors that cause impairment when they are higher, target samples to the right of the rightmost dashed black line (greater than the 75%-ile) receive a score of 1 (supporting evidence). For stressors that cause impairment when they are lower, samples to the left of the leftmost dashed black line

(less than the 25%-ile) receive a score of 1 (Table 3-2). In this example, stressor values for the target site (red lines) fall outside the range observed at better quality sites, strengthening the case for co-occurrence of
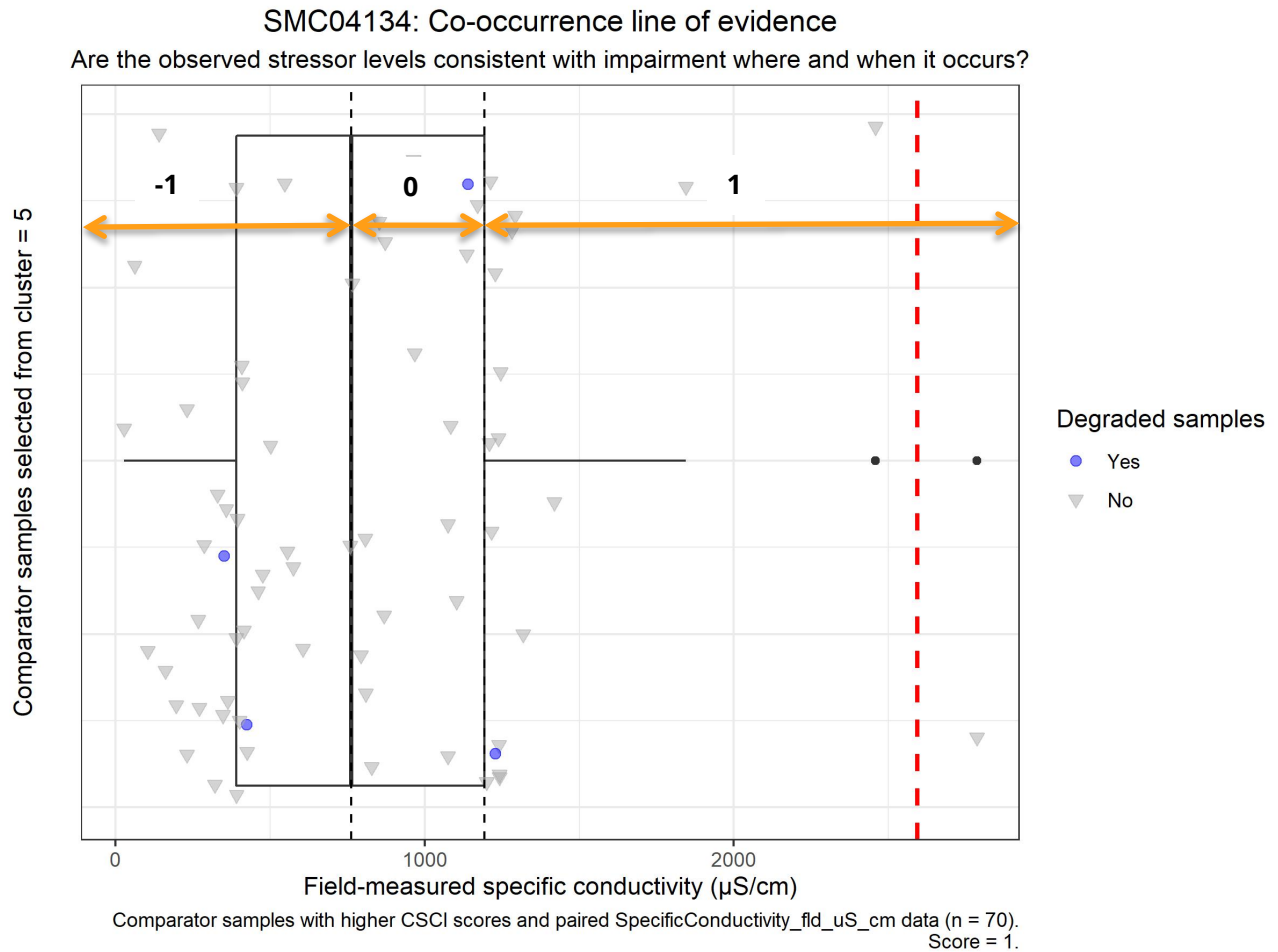
Figure 3-7. Example co-occurrence box plot created by `getCoOccur`. Arrows represent scoring regions. Numbers above the arrows reflect the score assigned in that region. This example reflects the case where increasing levels of the stressor increase impairment.

specific conductivity and impairment at the target site.

In addition to evaluating the co-occurrence line of evidence, this function also evaluates stressor-response relationships from inside the case. This evaluation produces a logistic regression (Figure 3-6). The logistic regression plots the Boolean indicator for degraded or not degraded against the stressor measurement for all comparator samples, where samples considered degraded are assigned a value of 1 and are represented by blue dots, and sites considered not degraded are assigned a value of zero and are represented as grey triangles. The logistic regression appears as a blue curve. The horizontal dashed lines represent the 50% and 20% probabilities of degraded condition for purposes of scoring. The red vertical dashed lines represent the stressor values for each sample obtained from the target site where there is also a paired biological sample. The score (shown in the lower right corner along with the number of samples) is determined from the location where the dashed red line crosses the blue curve. Both sets of graphics are combined into a single PDF file for ease of reference. The orange arrows and number show the zones for supporting (1), indeterminate (0), and refuting (-1) scores (Table 3-3).

SMC04134: Stressor-response (logistic regression) line of evidence

Are stressor levels sufficient to explain the observed impairment?

All comparator samples with CSCI and paired SpecificConductivity_fld_uS_cm data (n=161).
Score = 1.
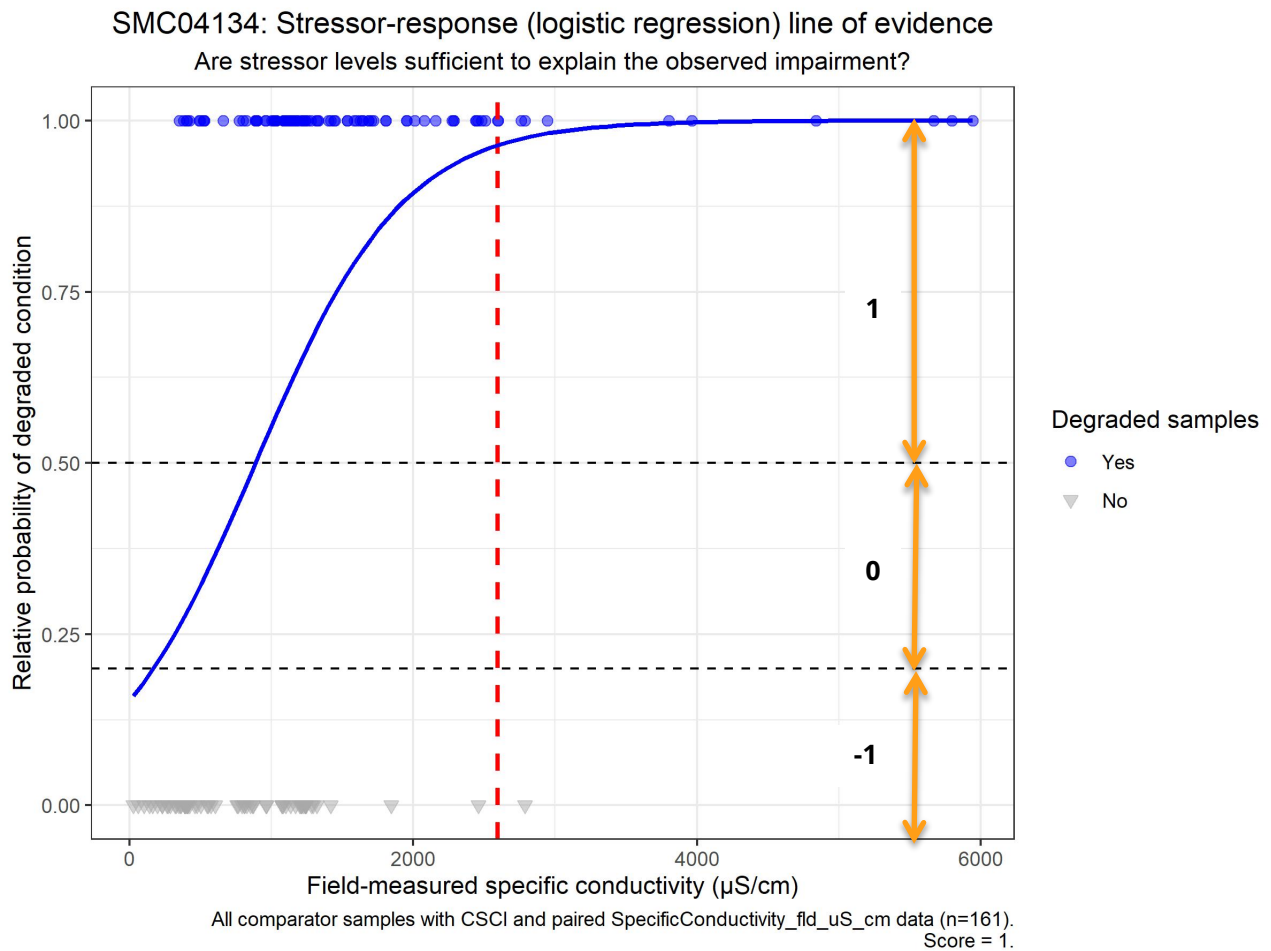
Figure 3-8. Example stressor-response logistic regression generated by `getCoOccur`.

The graphics are grouped by stressor, with the co-occurrence box plot appearing first and the logistic regression appearing second. In this example, one sample of percent fines in the target site occurs at a concentration associated with greater risk of degradation, supporting this as a potential cause. The final

output from this function is a text file containing information related to the scoring and the final scores for both the co-occurrence box plot and the stressor-response logistic regression.

Table 3-2. Scoring criteria for the co-occurrence line of evidence.

Direction of increased stress	Target site percentile	Score
Increase	>75th	+1
	<50th	-1
	50th ≤ observation ≤75th	0
Decrease	<25th	+1
	>50th	-1
	25th ≤ observation ≤50th	0

Table 3-3. Scoring criteria for the stressor-response from the stressor-response logistic regression line of evidence.

Target site percentile	Score
>50th	+1
<20th	-1
20th ≤ observation ≤50th	0

`getBioStressorResponses`. This function evaluates two separate lines of evidence: stressor-response relationships from the field (data from the case) and stressor-response relationships from other field studies (data from outside the case). To do so, it uses paired stressor and biological response data, where the responses can be either benthic macroinvertebrates or algae, as specified in an argument to the function.

This function generates linear regressions of the response on the stressor for each stressor-response combination. Two separate data sets are graphed in one image (Figure 3-7): the set of paired stressor-response samples from the comparator sites associated with the target site (data from the case), and the set of paired stressor-response samples from all the data in the dataset (data from elsewhere). If fewer than 20 paired stressor-response samples are available in the entire dataset, no evaluation occurs, and the data are noted as a data gap.

For each of the two regressions, the regression line is shown as a solid line, where teal represents the cluster regression and black represents the entire dataset regression. The 95% confidence intervals are shown by shaded bands of the same color as the regression line. The prediction intervals (defined by `predint`, set at 75% by default) are shown by dashed lines on either side of and in the same color as the regression lines. The individual sample data are shown as teal dots for comparator samples and as gray dots for samples from elsewhere. Blue outlines on dots indicate reference sites. Lastly, the target site data are shown as red triangles. The regression graphs also show the regression equations for the cluster

samples and all samples in the dataset on the lower right-hand side of the graph. Additionally, the three measures used to determine the scores for both lines of evidence are printed following the regression equations: the regression coefficient r^2 value, the statistical significance probability p-value of the slope, and sample size (n).
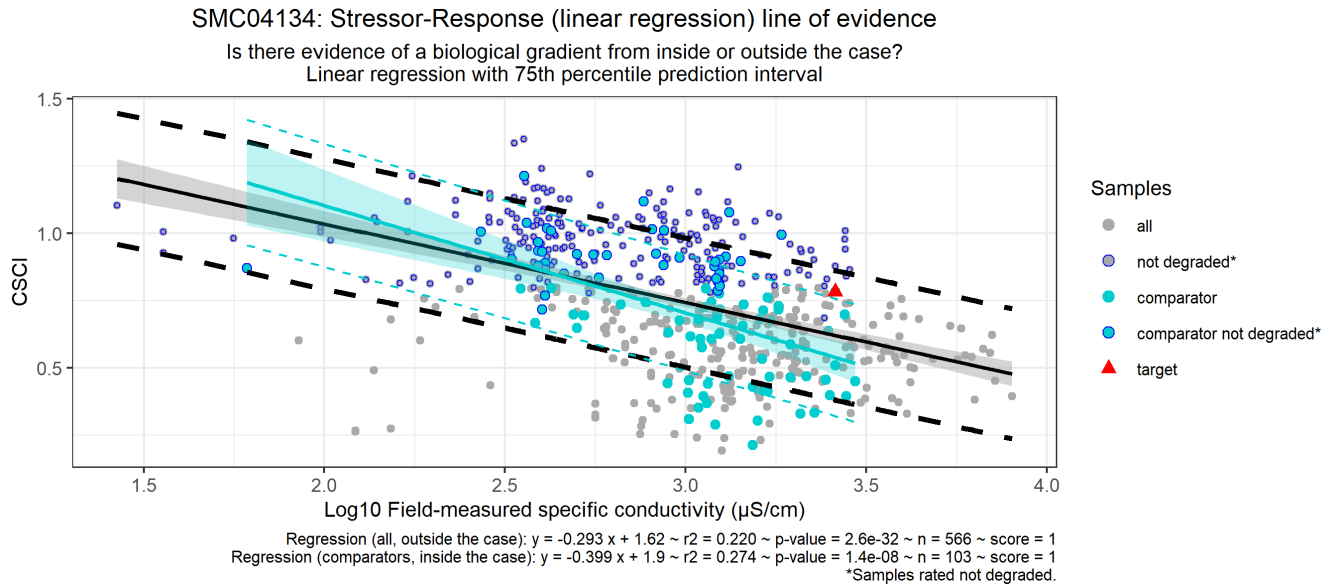
Figure 3-9. Example linear regression generated by `getBioStressorResponses`.

Each regression (comparator samples and all samples) is scored independently, as they represent different lines of evidence. Therefore, for scoring, the sample size for the regression must be at least 5, the p-value must be less than or equal to 0.1, and the r^2 value must be greater than or equal to 0.1. If these criteria are met and the slope is in the same direction as expected, then the score is +1 (the evidence supports the stressor as a cause of impairment). However, if the criteria are met and the slope is in the opposite direction from expected, then the score is -1 (the evidence refutes the stressor as a cause of impairment). In all other cases, the stressor receives a score of 0 (inconclusive). Table 3-3 presents a summary of the scoring method.

Table 3-4. Scoring criteria for the stressor-response from the field (linear regression) line of evidence.

Regression criteria met?	Slope in expected direction?	Score
Yes	Yes	+1
Yes	No	-1
No	Yes	0
No	Not evaluated	0

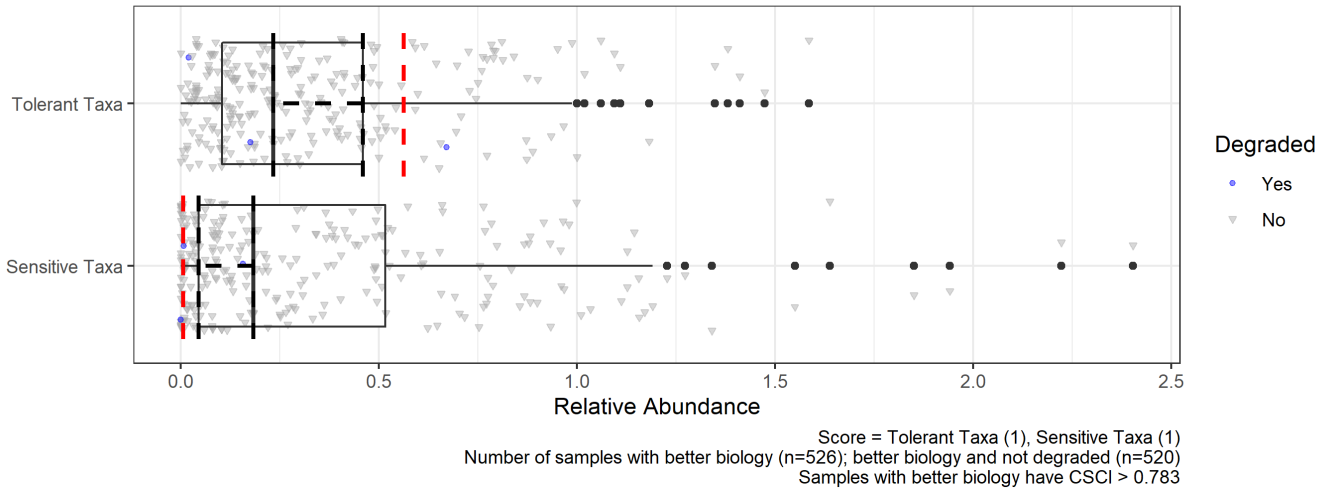Regression criteria are: n ≥ 5, r^2 ≥ 0.1, and p-value ≤ 0.1.

`getVerifiedPredictions.` Verified predictions are analyses that determine whether known diagnostic indicators for certain candidate stressors yield expected results for the stressor. Phrased as a hypothesis (prediction) this can be stated as "If the stressor affects stressor-sensitive taxa, there should be relatively fewer sensitive and relatively more tolerant taxa where the stressor is present." This function prepares box plots using the relative abundance of taxa sensitive or tolerant to a specific stressor, as determined using field-derived tolerance values for that stressor, for samples having better biological condition (i.e., higher CSCI score) than the target site. The box plots (Figure 3-9) show the distribution of relative abundance values of sensitive or tolerant taxa from all sites with better biological quality in the cluster. Sensitive and tolerant taxa are defined using the stressor-specific tolerance values derived from broad field observation data and specified within the input data structure. The white box of the box plot shows the interquartile range (IQR, the edges of the box), the median (the black line inside the box), and the horizontal lines to either side represent ±1.5 * IQR. Dots falling outside the horizontal lines represent outliers. The blue dots represent samples exhibiting better biology than the target samples. The red vertical dashed lines represent the relative abundance values for sensitive or tolerant taxa for each biological sample obtained from the target site. The number of samples represented, and the scores assigned to each target site sample, are shown in the bottom right corner of the graphic.

This function compares the relative abundance values for sensitive or tolerant taxa from the target site for a given stressor (e.g., conductivity) to the corresponding relative abundance values from all sites in the same cluster that have better biological response scores (CSCIs) than the worst biological response score observed in the target site, if multiple bioassessment scores are available for the target site (Table 3-4). If a target site relative abundance of tolerant taxa is greater than the 75th percentile of the relative abundance values of tolerant taxa reported for sites with better biology, or the target site relative abundance of sensitive taxa is less than the 25th percentile of the relative abundance values of sensitive taxa reported for sites with better biology, then the stressor to which taxa are either sensitive or tolerant receives a +1 score (supports the case for the stressor as a cause of impairment). If the target site relative abundance values are below (tolerant) or above (sensitive) the 50th percentile of stressor measurements reported for sites with better biology, then that stressor measurement receives a -1 score (refutes the case for the stressor as a cause of impairment). If the target site relative abundances are between the 50th and 75th percentile (tolerant) or the 25th and 50th percentile (sensitive) of relative abundances reported for sites with better biology, that stressor receives a 0 score (indeterminate).

SMC04134: Verified prediction line of evidence for Field-measured specific conductivity (μS/cm)

Do the data support the prediction that the abundance of sensitive taxa will be lower and tolerant taxa will be higher than observed at comparator sites with better biology?

Score = Tolerant Taxa (1), Sensitive Taxa (1)
Number of samples with better biology (n=526); better biology and not degraded (n=520)
Samples with better biology have CSCI > 0.783

Figure 3-10. Example box plot for verified prediction for fine sediment in the reach, generated by `getVerifiedPredictions`.

Table 3-5. Scoring criteria for the co-occurrence line of evidence.

Relative Abundance Type	Target Site percentile	Score
Tolerant taxa	>75th	+1
	<50th	-1
	50th ≤ observation ≤ 75th	0
Sensitive taxa	<25th	+1
	>50th	-1
	25th ≤ observation ≤ 50th	0

3.7 Evaluate Data from Elsewhere

Evaluating data from outside the case requires evaluating the data available from laboratory data, other literature, and field sites elsewhere. Using this additional information can strengthen the evidence for or against a particular candidate cause. Laboratory data such as Toxicity Identification Evaluations (TIEs) and other relevant studies published in the peer-reviewed or grey literature may be used by the causal assessor, but do not lend themselves to incorporation into an automated screening tool like the CASTool. The only line of evidence from outside the case that is evaluated as part of the CASTool includes Stressor-Response relationship from outside the case, which was addressed in Section 3.6 under the function getBioStressorResponses. The details around this evaluation are discussed thoroughly there and will not be repeated here.

4.0 Identify Probable Cause – Weight of Evidence Scoring

The final step in the causal assessment process is identifying a probable cause. A weight of evidence (WoE) scoring algorithm was developed by the Project team for integrating the results of different lines of evidence for a given test site. For any stressor for which data are available, the algorithm is designed to determine whether a stressor is unlikely, likely, or indeterminate (evidence is conflicting and further monitoring is needed). The CASTool uses the most-commonly evaluated lines of evidence for which quantitative data are available (Table 4-1).

Table 4-1. Example of weight of evidence scoring results for a biological site in the City of San Diego.

Sample Date	Stressor	% Rank	CO	SR Inside Log	SR Inside Lin	SR Outside Lin	# Lines of Evidence Inside Sup	# Lines of Evidence Inside Ref	# Lines of Evidence Inside Indet	# Lines of Evidence Inside NE	# Lines of Evidence Outside Sup	# Lines of Evidence Outside Ref	# Lines of Evidence Outside Indet	# Lines of Evidence Outside NE	# Lines of Evidence (total) Sup	# Lines of Evidence (total) Ref	# Lines of Evidence (total) Indet	# Lines of Evidence (total) NE	WoE
2012-05-23	Chloride (mg/L)	87.5	1	0	1	1	2	0	1	0	1	0	0	0	3	0	1	0	Sup
	Manganese (µg/L), dissolved	100	1	1	0	0	2	0	1	0	0	0	1	0	2	0	2	0	Sup
	Manganese (µg/L), total	88.9	1	0	1	0	2	0	1	0	0	0	1	0	2	0	2	0	Sup
	DOC (mg/L)	100	1	0	1	1	2	0	1	0	1	0	0	0	3	0	1	0	Sup
	Silica (mg/L)	100	1	0	0	0	1	0	2	0	0	0	1	0	1	0	3	0	Indet
	TN (mg/L)	92.9	1	0	0	1	1	0	2	0	1	0	0	0	2	0	2	0	Sup
	TDS (mg/L)	83.3	-1	0	1	1	1	1	1	0	1	0	0	0	2	1	1	0	Sup

Key:
CO = Co-occurrence
SR = Stressor-response
WoE = Weight of Evidence

Log = logistic regression
Lin = linear regresson

Sup = supports
Ref = refutes
Indet = indeterminate
NE = not evaluated

As a screening tool, the CASTool provides summarized information for each of the potential stressors and the lines of evidence evaluated, which are intended to assist the user in drawing reasonable conclusions about causes of impairment (or vulnerability) at a target site. The scoring methods described previously are combined and summarized using the final interpretive function of the tool which acquires the full weight of evidence for each stressor evaluated.

The Weight of Evidence function (getWoE) combines the results of the evaluations performed for each potential stressor by line of evidence for each sample exhibiting impaired or degraded biology. The evidence is presented as supporting, refuting, indeterminate, or not evaluated, and are presented as individual lines of evidence, inside/outside the case evidence, or overall evidence, with the biological index representing the observed response. Additionally, the weight of evidence is summarized by stressor, over all

samples for which that stressor was evaluated. Lastly, the function provides a summary of the linear regression results for stressor-response inside or outside the case for which biological metrics represent the observed response.

This function generates several data structures summarizing the weight of evidence for or against a potential stressor as a cause of impairment at the target site. The first and most detailed text (`TargetSite_ResponseCommunity`_WoE_DetailedLoEs.tab) presents scores for each stressor in a single sample that was paired with an impaired biological response index for each line of evidence, where the biological response is the index for the indicated biological community. Valid scores are +1, 0, and -1, representing supporting, indeterminate, or refuting evidence for the stressor as a cause of impairment. The verified prediction line of evidence is comprised of two separate evaluations (sensitive taxa and tolerant taxa) presented separately. A value of NE indicates that the line of evidence was not evaluated for that stressor. A separate, but related file (`TargetSite_ResponseCommunity`_WoE_DetailedMetricsLoEs.tab) presents the same results for individual metrics that comprise the index.

A third file, `TargetSite_ResponseCommunity`_WoE_ScoresTable.tab, summarizes the number of lines of evidence inside or outside the case that support, refute, provide indeterminate evidence, or were not evaluated, and the total number of lines of evidence supporting, refuting, providing indeterminate evidence, or not evaluated. Finally, this file output concludes with a statement of whether the overall weight of evidence for a stressor in a particular sample supports, refutes, or is indeterminate as a cause of biological impairment.

The final output file, `TargetSite_ResponseCommunity`_WoE_ExecSummary.tab, summarizes the stressors as a stressor group, with the combined weights of evidence by line of evidence, and the overall weight of evidence across lines of evidence, also by stressor group.

4.1 Summarize the Analysis

The final site-specific function provided in the CASTool package (`getReport)` creates a report of the findings including the date and time of the analysis, the user who ran the analysis, user-specified inputs, general site information, results by response community, and data gaps. This function uses a template to combine data files generated from other functions in the CASTool package. The report is generated as an HTML file using an R Markdown document template (rmd file) that can be readily viewed in any browser.

5.0 Shiny Web Application of the CASTool

The CASTool uses a pull down/auto-complete list of all known sites site for siteinput and modular construction of the software to readily accommodate future enhancements. There is a map component for checking site locations or discovering sites in a known area. The CASTool analyzes the available data for a site, which are often incomplete in terms of stressor data or may have different relationships between biological and stressor data for different time periods. As the objective of the SEP is to develop a screening causal assessment approach, it is expected that follow-up sampling and additional causal assessment analyses will be needed prior to identifying restoration potential and appropriate best management practices.

For the CASTool, the only input is the Station ID. The user selects the Station ID from the pull-down list and then presses the "Calculate Results" button. Once results are finalized the "Download Results" button is enabled. The download is a zip file of all result files. During calculation a progress bar and popup is shown in the lower right corner, and any warnings or messages are displayed on screen on the "Console" tab.

6.0 CASTool Output Files and Directory Structure

Files generated by each of the functions in the CASTool package were described along with the functions. This section focuses on the output organization, so that the analyst can quickly identify the desired graphics or tables desired.

If the user runs the CASTool through the shiny interface, a compressed file is made available for download. If the user runs the CASTool through the console, the user must identify a location on their hard drive or network for the output files to be written. In both cases, the structure of the output files is similar. The root directory is called "Results." The target site identifiers are used as the names for each site's results output. General information falls within the target site folder, and results that are more specific to the biological community analysis are placed into folders named with the biological community type. The hierarchical representation is shown in Figure 6-1.

The figure shows only a high-level view of the directory structure. Each box, unless it specifies "file," represents a folder or directory. A variety of results files, mostly graphics, will be stored in each directory based on the function that generated the data. Therefore, SiteInfo contains general information about the target site; CandidateCauses contains the stressor boxplots from which potential candidate causes are selected; etc. The final results summary report is found in the site folder as an html (browser-readable) file. More detailed supporting evidence appears in each biological community's folder. The only folders that are written are those that are supported by the data. In other words, if there are no matching stressor/response data, no line of evidence folders will be written. If no algal data exist, the ALGAE folder won't be written.
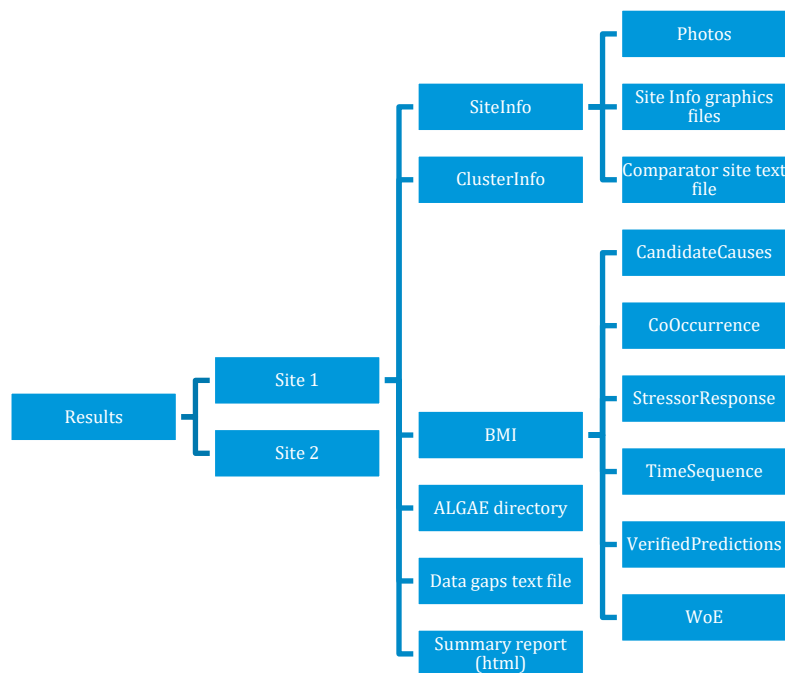
Figure 6-1. Hierarchical output structure showing folders written by the CASTool, each of which contain supporting graphics and text files specific to the folder's name.

7.0 References

Mazor, R., Ode, P. R., Rehn, A. C., et al. (2015). The California Stream Condition Index (CSCI): Interim instructions for calculating scores using GIS and R (SCCWRP Technical Report #883/SWAMP-SOP-2015-0004). Retrieved from Costa Mesa, CA: http://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/ docs/csci_scoring_instruct.pdf.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Tetra Tech. 2019. Preliminary Restoration and Protection Prioritization (RPP) Framework for the Supplemental Environmental Project (SEP) Support Restoring and Protecting Our Waterways: Bio-assessment Tools and Priorities for Action. Prepared for City of San Diego by Tetra Tech, Inc.

U.S. EPA (Environmental Protection Agency). Causal Analysis/Diagnosis Decision Information System (CADDIS): *Insert Web Page Title*. Retrieved: *Insert Month, Day, Year*. Available from: *www.epa.gov/caddis*. Office of Research and Development, Washington, DC, 2017. Accessed June 11, 2019.

Wickham et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Appendix A-1. Data used in clustering stream reaches

Data from the NHDPlus dataset and StreamCat were used to cluster stream reaches, as described in Section 2. For each reach dataset (high and low elevation), Slope and stream order values used to indicate unknown values (-9998 and -9, respectively) were replaced with "NA." Stream types identified as coastline and pipeline were removed from the dataset, as they do not represent regulated freshwater stream waterbodies. StreamCat variables associated with human alteration such as atmospheric deposition or land cover were removed from the dataset. Lastly, StreamCat variables representing natural characteristics for which at least 75% of stream reaches had no data were removed from the dataset. Lognormally distributed variables were log-transformed and percentage data were square root transformed to better approximate normally distributed data. Table A-1 lists the NHDPlus and StreamCat variables used in the final clustering algorithm, along with the transformations applied.

Table A-1. Variables used to identify comparator sites

Variable	Definition	Category	Source	Transformation
COMID	Common identifier of the NHD feature	Information	NHD+	none
Al2O3Cat	Mean % of lithological aluminum oxide (Al2O3) content in surface or near surface geology within catchment	Lithology	StreamCat	none
Al2O3Ws	Mean % of lithological aluminum oxide (Al2O3) content in surface or near surface geology within watershed	Lithology	StreamCat	none
BFICat	Base flow is the component of streamflow that can be attributed to ground-water discharge into streams. The BFI is the ratio of base flow to total flow, expressed as a percentage, within catchment	Hydrology	StreamCat	none
BFIWs	Base flow is the component of streamflow that can be attributed to ground-water discharge into streams. The BFI is the ratio of base flow to total flow, expressed as a percentage, within watershed	Hydrology	StreamCat	none
CaOCat	Mean % of lithological calcium oxide (CaO) content in surface or near surface geology within catchment	Lithology	StreamCat	logarithmic
CaOWs	Mean % of lithological calcium oxide (CaO) content in surface or near surface geology within watershed	Lithology	StreamCat	logarithmic
CatAreaSqKm	Area of local NHDPlus catchment (square km)	Topography	StreamCat	logarithmic

Variable	Definition	Category	Source	Transformation
CatAreaSqKmRp100	Area of local NHDPlus catchments (square km) within a 100-m buffer of NHD streams.	Topography	StreamCat	logarithmic
ClayCat	Mean % clay content of soils (STATSGO) within catchment	Soils	StreamCat	logarithmic
ClayWs	Mean % clay content of soils (STATSGO) within watershed	Soils	StreamCat	logarithmic
CompStrgthCat	Mean lithological uniaxial compressive strength (megaPascals) content in surface or near surface geology within catchment	Lithology	StreamCat	none
CompStrgthWs	Mean lithological uniaxial compressive strength (megaPascals) content in surface or near surface geology within watershed	Lithology	StreamCat	none
ElevCat	Mean catchment elevation (m)	Topography	StreamCat	square root
ElevWs	Mean watershed elevation (m)	Topography	StreamCat	square root
Fe2O3Cat	Mean % of lithological ferric oxide (Fe2O3) content in surface or near surface geology within catchment	Lithology	StreamCat	logarithmic
Fe2O3Ws	Mean % of lithological ferric oxide (Fe2O3) content in surface or near surface geology within watershed	Lithology	StreamCat	logarithmic
HydrlCondCat	Mean lithological hydraulic conductivity (micrometers per second) content in surface or near surface geology within catchment	Lithology	StreamCat	logarithmic
HydrlCondWs	Mean lithological hydraulic conductivity (micrometers per second) content in surface or near surface geology within watershed	Lithology	StreamCat	logarithmic
K2OCat	Mean % of lithological potassium oxide (K2O) content in surface or near surface geology within catchment	Lithology	StreamCat	logarithmic
K2OWs	Mean % of lithological potassium oxide (K2O) content in surface or near surface geology within watershed	Lithology	StreamCat	logarithmic
KffactCat	Mean soil erodibility (Kf) factor (unitless) of soils within catchment. The Kffactor is used in the Universal Soil Loss Equation (USLE) and represents a relative index of susceptibility of bare, cultivated soil to particle detachment and transport by rainfall.	Soils	StreamCat	none
KffactWs	Mean soil erodibility (Kf) factor (unitless) of soils within watershed. The Kffactor is used in the Universal Soil Loss Equation (USLE) and represents a relative index of susceptibility of	Soils	StreamCat	none

Variable	Definition	Category	Source	Transformation
	bare, cultivated soil to particle detachment and transport by rainfall.			
MgOCat	Mean % of lithological magnesium oxide (MgO) content in surface or near surface geology within catchment	Lithology	StreamCat	logarithmic
MgOWs	Mean % of lithological magnesium oxide (MgO) content in surface or near surface geology within watershed	Lithology	StreamCat	logarithmic
Na2OCat	Mean % of lithological sodium oxide (Na2O) content in surface or near surface geology within catchment	Lithology	StreamCat	none
Na2OWs	Mean % of lithological sodium oxide (Na2O) content in surface or near surface geology within watershed	Lithology	StreamCat	none
NCat	Mean % of lithological nitrogen (N) content in surface or near surface geology within catchment	Lithology	StreamCat	square root
NWs	Mean % of lithological nitrogen (N) content in surface or near surface geology within watershed	Lithology	StreamCat	square root
OmCat	Mean organic matter content (% by weight) of soils (STATSGO) within catchment	Soils	StreamCat	logarithmic
OmWs	Mean organic matter content (% by weight) of soils (STATSGO) within watershed	Soils	StreamCat	logarithmic
P2O5Cat	Mean % of lithological phosphorous oxide (P2O5) content in surface or near surface geology within catchment	Lithology	StreamCat	square root
P2O5Ws	Mean % of lithological phosphorous oxide (P2O5) content in surface or near surface geology within watershed	Lithology	StreamCat	square root
PctAlluvCoastCat	% of catchment area classified as lithology type: alluvium and fine-textured coastal zone sediment	Lithology	StreamCat	logarithmic
PctAlluvCoastWs	% of watershed area classified as lithology type: alluvium and fine-textured coastal zone sediment	Lithology	StreamCat	logarithmic
PctNonCarbResidCat	% of catchment area classified as lithology type: non-carbonate residual material	Lithology	StreamCat	logarithmic
PctNonCarbResidWs	% of watershed area classified as lithology type: non-carbonate residual material	Lithology	StreamCat	logarithmic
PctSilicicCat	% of catchment area classified as lithology type: silicic residual material	Lithology	StreamCat	none

Variable	Definition	Category	Source	Transformation
PctSilicicWs	% of watershed area classified as lithology type: silicic residual material	Lithology	StreamCat	none
PermCat	Mean permeability (cm/hour) of soils (STATSGO) within catchment	Soils	StreamCat	logarithmic
PermWs	Mean permeability (cm/hour) of soils (STATSGO) within watershed	Soils	StreamCat	logarithmic
Precip08Cat	PRISM climate data - Mean precipitation (mm) within the catchment. Period: 2008	Climate	StreamCat	logarithmic
Precip08Ws	PRISM climate data - Mean precipitation (mm) within the watershed. Period: 2008	Climate	StreamCat	logarithmic
Precip09Cat	PRISM climate data - Mean precipitation (mm) within the catchment. Period: 2009	Climate	StreamCat	logarithmic
Precip09Ws	PRISM climate data - Mean precipitation (mm) within the watershed. Period: 2009	Climate	StreamCat	logarithmic
Precip8110Cat	PRISM climate data - 30-year normal mean precipitation (mm): Annual period: 1981-2010 within the catchment	Climate	StreamCat	logarithmic
Precip8110Ws	PRISM climate data - 30-year normal mean precipitation (mm): Annual period: 1981-2010 within the watershed	Climate	StreamCat	logarithmic
QC_01	Mean January flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_02	Mean February flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_03	Mean March flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_04	Mean April flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_05	Mean May flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_06	Mean June flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_07	Mean July flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic

Variable	Definition	Category	Source	Transformation
QC_08	Mean August flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_09	Mean September flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_10	Mean October flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_11	Mean November flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_12	Mean December flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
QC_MA	Mean annual flow with reference gage regression applied to QC (cfs). Best EROM estimate of "natural" mean flow.	Hydrology	NHD+	logarithmic
RckDepCat	Mean depth (cm) to bedrock of soils (STATSGO) within catchment	Soils	StreamCat	logarithmic
RckDepWs	Mean depth (cm) to bedrock of soils (STATSGO) within watershed	Soils	StreamCat	logarithmic
RunoffCat	Mean runoff (mm) within catchment	Hydrology	StreamCat	logarithmic
RunoffWs	Mean runoff (mm) within watershed	Hydrology	StreamCat	logarithmic
SandCat	Mean % sand content of soils (STATSGO) within catchment	Soils	StreamCat	none
SandWs	Mean % sand content of soils (STATSGO) within watershed	Soils	StreamCat	none
SCat	Mean % of lithological sulfur (S) content in surface or near surface geology within catchment	Lithology	StreamCat	square root
SiO2Cat	Mean % of lithological silicon dioxide (SiO2) content in surface or near surface geology within catchment	Lithology	StreamCat	none
SiO2Ws	Mean % of lithological silicon dioxide (SiO2) content in surface or near surface geology within watershed	Lithology	StreamCat	none
SLOPE	Slope of flowline (meters/meters) based on smoothed elevations; a value of -9998 means that no slope value is available. See Appendix A, step 22 for information about slope computation.	Topography	NHD+	square root

Variable	Definition	Category	Source	Transformation
SWs	Mean % of lithological sulfur (S) content in surface or near surface geology within watershed	Lithology	StreamCat	square root
Tmax8110Cat	PRISM climate data - 30-year normal maximum temperature (C°): Annual period: 1981-2010 within the catchment	Climate	StreamCat	none
Tmax8110Ws	PRISM climate data - 30-year normal maximum temperature (C°): Annual period: 1981-2010 within the watershed	Climate	StreamCat	none
Tmean08Cat	PRISM climate data - Mean temperature (C°)within the catchment. Period: 2008	Climate	StreamCat	none
Tmean08Ws	PRISM climate data - Mean temperature (C°)within the watershed. Period: 2008	Climate	StreamCat	none
Tmean09Cat	PRISM climate data - Mean temperature (C°)within the catchment. Period: 2009	Climate	StreamCat	none
Tmean09Ws	PRISM climate data - Mean temperature (C°)within the watershed. Period: 2009	Climate	StreamCat	none
Tmean8110Cat	PRISM climate data - 30-year normal mean temperature (C°): Annual period: 1981-2010 within the catchment	Climate	StreamCat	none
Tmean8110Ws	PRISM climate data - 30-year normal mean temperature (C°): Annual period: 1981-2010 within the watershed	Climate	StreamCat	none
Tmin8110Cat	PRISM climate data - 30-year normal minimum temperature (C°): Annual period: 1981-2010 within the catchment	Climate	StreamCat	none
Tmin8110Ws	PRISM climate data - 30-year normal minimum temperature (C°): Annual period: 1981-2010 within the watershed	Climate	StreamCat	none
WetIndexCat	Mean Composite Topographic Index (CTI)[Wetness Index] within catchment	Wetness	StreamCat	logarithmic
WetIndexWs	Mean Composite Topographic Index (CTI)[Wetness Index] within watershed	Wetness	StreamCat	logarithmic
WsAreaSqKm	Watershed area (square km) at NHDPlus stream segment outlet, i.e., at the most downstream location of the vector line segment	Topography	StreamCat	logarithmic
WsAreaSqKmRp100	Watershed area (square km) within a 100-m buffer of NHD streams	Topography	StreamCat	logarithmic
WtDepCat	Mean seasonal water table depth (cm) of soils (STATSGO) within catchment	Soils; baseflow	StreamCat	none
WtDepWs	Mean seasonal water table depth (cm) of soils (STATSGO) within watershed	Soils; baseflow	StreamCat	none

Appendix B. RPPTool Technical Memo

MEMO

To: Vicki Kalkirtz, Ruth Kolb (City of San Diego)

From: Clint Boschen, Jerry Diamond, Ann Roseberry Lincoln, Erik Leppo (Tetra Tech)

Subject: Restoration and Protection Prioritization (RPP) Framework for the Supplemental Environmental Project (SEP). Restoring and Protecting Our Waterways: Bio-assessment Tools and Priorities for Action

Table of Contents

Figures

Tables

Acronyms and Abbreviations

BCG	Biological Condition Gradient
CalFire	California Department of Forestry and Fire Protection, Fire and Resource Assessment Program
CASTool	Causal Assessment Screening Tool
COMID	Common identifier for each reach in the NHD
CSCI	California Stream Condition Index
ESA	Environmentally Sensitive Area
HUC	Hydrologic Unit Code
IPI	Index of Physical Integrity
km	kilometer
MHPA	Multi-habitat Planning Area
MSCP	Multiple Species Conservation Program
NASVI	Native Aquatic Species Vulnerability Index
NHD	National Hydrography Dataset
qt05	Fifth percentile
qt50	Median (fiftieth percentile)
qt95	Ninety-fifth percentile
RPPTool	Restoration and Protection Prioritization Tool
RPS	U.S. EPA's Recovery Potential Screening Tool
SANDAG	San Diego Association of Governments
SanGIS	SANDAG Regional GIS Data Warehouse Open Data Portal
SCAPE	Stream Classification and Priority Explorer
SCCWRP	Southern California Coastal Water Research Program
SEP	Supplemental Environmental Project
SMC	Stormwater Monitoring Coalition

1.0 Introduction

The overarching goal of the *Supplemental Environmental Project (SEP) Support Restoring and Protecting Our Waterways: Bio-assessment Tools and Priorities for Action* is to provide a user-friendly screening tool to prioritize streams in the City of San Diego (City) region of California for restoration and protection actions. The Causal Assessment Screening Tool (CASTool) previously developed under this SEP provides results that indicate which stressor(s) are likely causes of biological impairment (Tetra Tech (2019), CASTool; Appendix A). The Restoration and Protection Prioritization Tool (RPPTool), a companion to the CASTool in the SEP Bioassessment Toolkit, provides a screening application for decision makers to assign priorities to stream segments for restoration or protection. Separate restoration and protection RPP index scores are derived for each stream reach using multiple data sources, with opportunities for users to adjust the scoring weights of certain component metrics.

The RPPTool framework incorporates stressors identified as likely causes of biological impairment by the CASTool, analyses of recovery potential factors, threats, and co-benefit opportunities in a systematic way to prioritize water bodies for restoration or protection. A previous literature review as part of this SEP indicated that many restoration or protection frameworks currently used by states and local agencies are based on subjective input or are tailored for a specific objective (e.g., target streams to reduce nutrient loading). The approach developed in this SEP builds on US EPA's Recovery Potential Screening Tool (RPS) using regional, objective metrics to provide a systematic screening approach that can address multiple objectives in relation to overall ecological and multi-benefit improvements. The RPPTool provides a transparent framework for water resource managers, which can help support planning and decision making as well as provide a platform for discussions with regulators and stakeholders. The design and methods presented here outline the data incorporated into the tool, how they are used in the final scoring approach, and the user experience.

2.0 Background

The City of San Diego, and potentially other areas of California, have a continuing need to assign management priorities to their waterways. A variety of tools have been developed across the nation to identify priorities for restoration. Often such prioritization is conducted in an informal manner, based on subjective input from water managers and stakeholders. However, some efforts have used a more systematic approach. The project team reviewed several systematized approaches in developing the RPP including the following:

- U.S. EPA's Recovery Potential Screening Tool (RPS)
- Minnesota's Statewide Protection Prioritization for Streams
- Midwest Biodiversity Institute's Integrated Prioritization System (IPS)

In addition, we reviewed the following other restoration and protection prioritization efforts that have been completed for specific watersheds and regions. Most of these efforts used a qualitative approach, rather than an *a priori* scoring algorithm, in a region where the need for some form of restoration was already determined.

- Cuyahoga River (Ohio) GIS-based Model
- Toronto and Region Conservation Authority Integrated Restoration Prioritization Approach
- North Atlantic Aquatic Connectivity Collaborative Customizable Prioritization Tool
- Yellow River Watershed (Alabama and Florida) Prioritized Restoration Sites
- Danish Stream Fauna Index
- Santa Clara River (California) Riparian Conservation Prioritization
- Total Maximum Daily Load (TMDL) Prioritization – Ecological Function and Services Approach
- Philadelphia Metropolitan Area Stream Restoration Prioritization Schema

A common theme from the review was the importance of understanding the goal of restoration or protection efforts to guide the prioritization approach. For example, if an entity wanted to restore as many individual water bodies as possible, they would likely target water bodies that were only marginally impaired, or for which minimal effort would lead to restoration. On the other hand, if an entity wanted to restore ecological integrity to a larger watershed, they could target water bodies that serve as corridors or links within the watershed in order to achieve the greatest improvement of water bodies in the watershed.

Our review of these existing approaches led to the structure and function of the RPPTool as described in Figure 2-1-1. The RPPTool is designed to provide the City of San Diego with a transparent and effective tool for setting restoration and protection priorities in their waterways, given the available data.

The RPP Index is comprised of three subindices—Potential, Threats, and Opportunities—that combine to produce restoration and protection index scores independently for each reach assessed with the tool, as follows:

- Restoration Index Score – the ecological uplift that could be attained through restoration of the reach, resulting in significant improvement of the biological community from one biological condition class to a higher class, incorporating likely threats and opportunities that would affect restoration success.

- Protection Index Score – the relatively high ecological condition of the reach that can be preserved through protection, that is, prevention of degradation of biological community from one biological condition class to a lower class, while accounting for threats and opportunities that would affect protection success.

Much like U.S. EPA's RPS Tool, the results of Potential, Threats, and Opportunities subindices are scored individually and then combined to form the overall RPP index scores. This tool, while similar to the RPS tool, focuses on local or regional data, as opposed to national data and calculates indicators based on the catchment scale, as opposed to the watershed scale. Moreover, the purpose of the RPPTool and the SEP overall is to develop a useful restoration and protection prioritization approach based largely on ecological factors and including social co-benefits where those co-benefits will maximize the ecological condition. Non-ecological factors such as cost, feasibility, and accessibility are not considered in the RPPTool approach. However, use of the RPPTool, combined with other tools used by the City to target management efforts, can provide a powerful approach for achieving meaningful improvements in aquatic life.

The purpose and data inputs of each subindex are described in Chapter 3. In Chapter 4, we describe the RPPTool user process and scoring methods. Chapter 5 presents a discussion of the uses of RPP, including examples of data inputs and scoring results.

Figure 2-1-1. Structure of the RPP index with three subindices and nine component indicators.

3.0 RPPTool Data

The core data used in the RPPTool are comprised of geographic data, such as the natural stream network as mapped by the U.S. Geological Survey (USGS) which forms the analytical skeleton for which the indicators are calculated; observed or measured data, such as the measured biological condition based on samples obtained of stream macroinvertebrates and algae; and predicted data, such as the predicted biological condition based on a model of landscape constraints. Measured and predicted data are used to calculate the indices. These data are described in the following sections.

3.1 Stream Network

The stream reach is the basic unit for evaluation in the RPPTool. Stream reaches for the Stormwater Monitoring Coalition (SMC) region were obtained from the National Hydrography Dataset (NHD), version 2, NHDFlowline and NHDPlusCatchment shapefiles and the PlusFlowlineVAA table (USGS and USEPA, 2012). The regional boundary was provided by Southern California Coastal Water Research Program (SCCWRP) (Gillett, 2018). To create a shapefile containing only the reach files within the SMC boundary without truncating their length or catchment, any reach that intersected or was contained within the boundary was included in the clipped extent. This process ensured that reach catchments extending beyond the boundary were retained intact and that no catchment area for reaches not within the region were included.

The NHDFlowline shapefile provides a common identifier ("COMID") and length ("LengthKm") for each flowline (or stream reach in RPPTool terminology). The NHDCatchment supplies the catchment area (AreaSqKm) for each catchment associated with the stream reach. Lastly, the Value Added Attributes (PlusFlowlineVAA) table joins the shapefile data (flowlines and catchments) with other relevant attributes such as FromNode (the point at the upstream end of the reach which connects the reach to one or more reaches upstream), ToNode (the point at the downstream end of the reach), and StartFlag (indicating whether the reach is a headwater reach). Combined, these attributes are used to identify the connected reaches for both connectivity indicators (biological condition and stressor) and headwater status if a headwater bonus will be applied.

3.2 Observed Biological Condition

The California Stream Condition Index (CSCI) is a biological scoring tool that helps aquatic resource managers translate complex data about benthic macroinvertebrates found living in a stream into an overall measure of stream health (Mazor et al., 2016). The CSCI score indicates whether, and to what degree, the ecology of a stream is altered from a healthy state. Direct measures of ecosystem health like the CSCI are preferable to those based on chemical or physical measurements for many management questions. Living organisms integrate the effects of multiple stressors, such as sedimentation, nutrient enrichment and riparian disturbance, over both space and time.

The CSCI score is a measure of how well a site's observed condition matches its predicted, or expected, condition. Expected values of a set of ecological indicators are predicted using statistical models. Predictions

are based on natural environmental variables resulting in a site-specific prediction for each site; greater deviations from this expectation indicate a greater likelihood of degradation. The CSCI score is calculated by comparing the expected condition with actual (observed) results. CSCI scores range from 0 (highly degraded) to greater than 1 (equivalent to reference (Ode et al., 2016).

Observed biological data and calculated metrics, including the final CSCI score, are obtained from the SMC data portal and are used in both the CASTool and the RPPTool. Currently the RPPTool uses all the data available in the SMC portal, which has biological community sample data ranging from 1996 to 2017 (Figure 3-1, SCCWRP Staff (2019)). The user may limit the range of data desired to be used. The default range is from the present back twelve years (currently 2008-2020).

3.3 Predicted Biological Condition

The Stream Classification and Priority Explorer (SCAPE), developed by SCCWRP and its partners, provides a visualization for data that predict the degree to which stream biological integrity scores are likely to be limited, or "constrained," by urban and agricultural development (Beck et al., 2019). The SCAPE tool utilizes a statewide landscape model to predict ranges of likely CSCI scores. Additionally, the SCAPE tool uses the CSCI threshold of 0.79 as the biological threshold for decision-making. The RPPTool utilizes the SCAPE CSCI predictions for each reach in the SMC region to evaluate the potential for biological uplift or protection from degradation that a reach might experience if restored or protected.
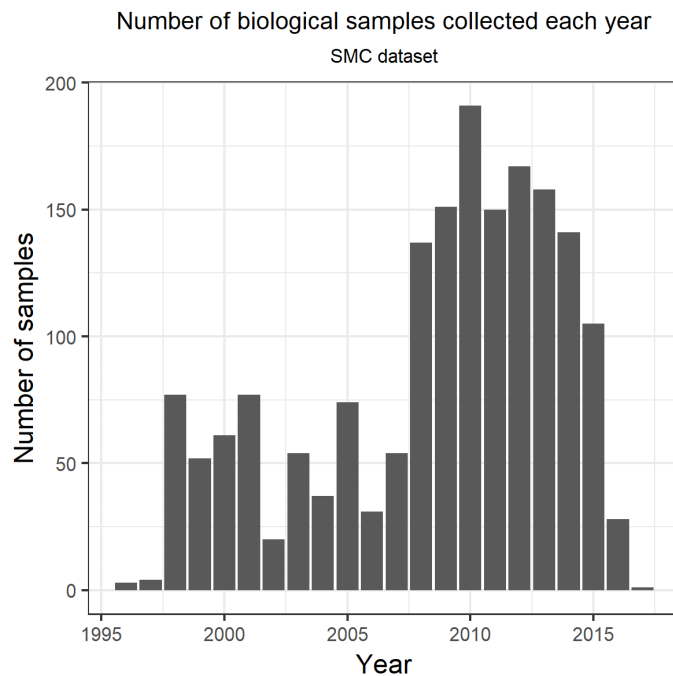
Figure 3-1. Number of biological samples in the SMC dataset by year between 2000 and 2017.

3.4 Biological Condition Gradient

The Biological Condition Gradient (BCG) is a scientific framework that can be used by states, tribes, territories and counties to interpret biological responses from the cumulative effects of stressors for all types of water bodies (USEPA, 2016). The BCG was proposed in 2006 as a conceptual framework developed by U.S. EPA in partnership with scientists from states, USGS, and the academic community. The BCG builds upon and complements previous biological criteria approaches to provide a more refined and detailed measure of biological condition using up to six levels shown below (Figure 3-2).

To better categorize the degree of anticipated uplift or protection, the RPPTool overlays the BCG tiers onto the CSCI observations and predictions (Figure 3-3). This has the added benefit of not relying on a single biological threshold for decision-making but a continuous scale that can indicate whether substantial biological improvement is possible even if the projected lift is still below the CSCI threshold of 0.79. Similarly, the projected benefit from added protection can be better categorized.
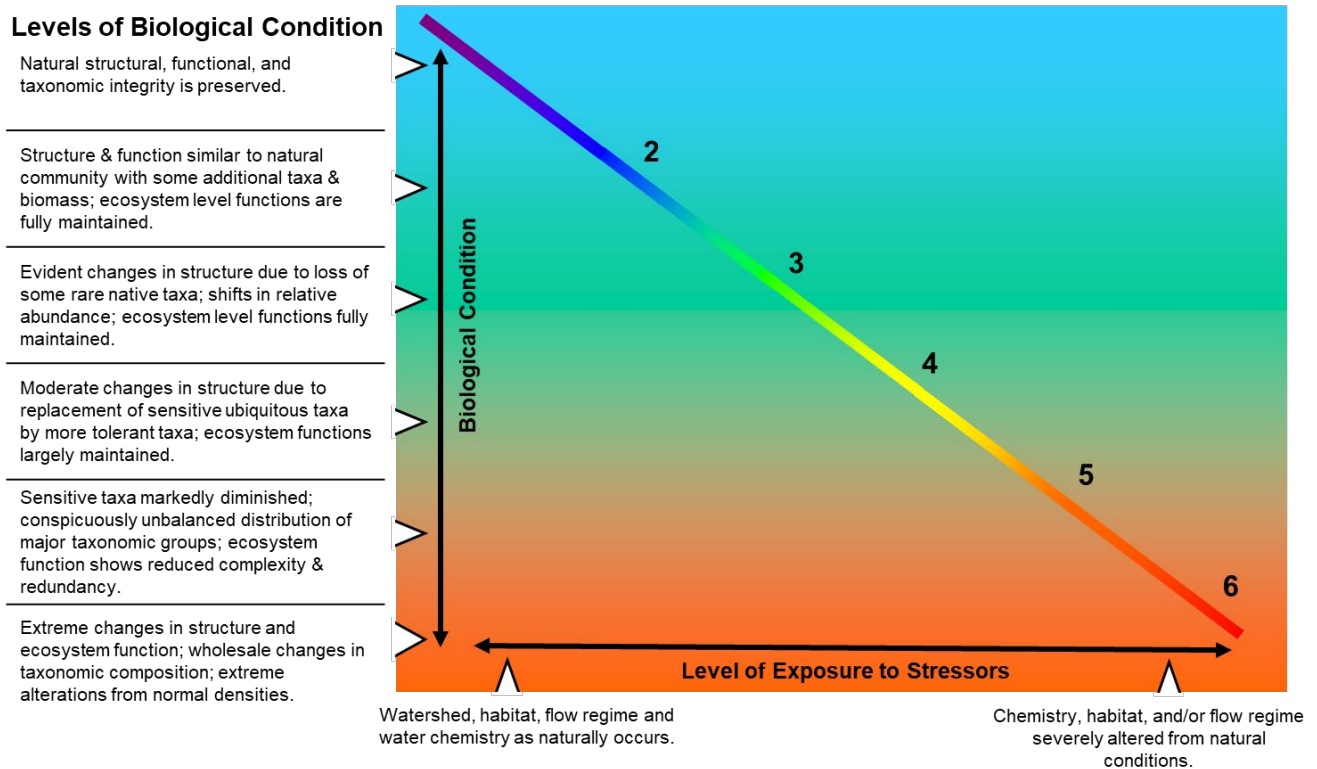
Figure 3-2. Biological Condition Gradient with six tiers of condition related to stressor exposure. (USEPA, 2016)
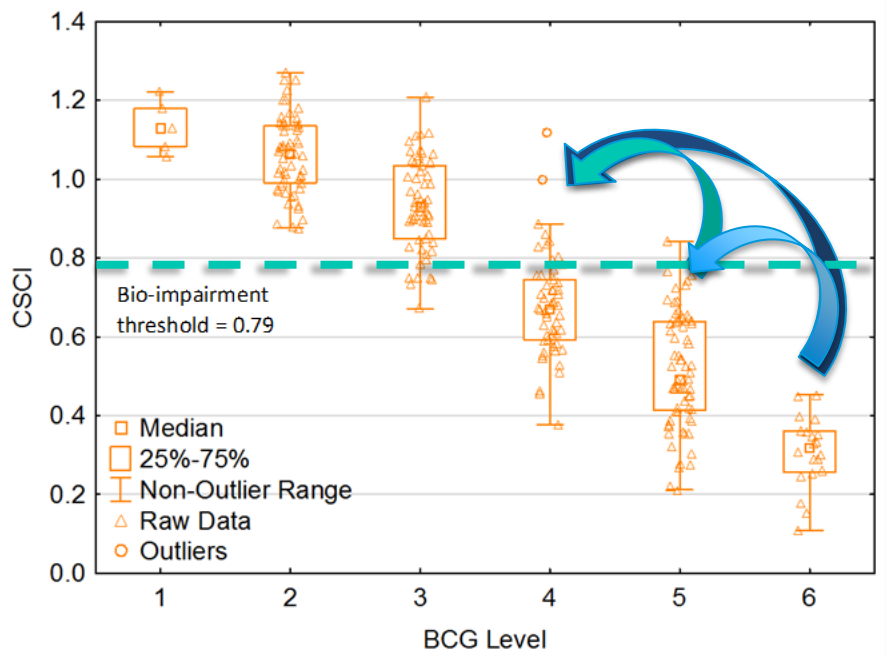
Figure 3-3. Example BCG tiers versus CSCI score with arrows indicating potential uplift.

3.5 Stressors

The CASTool was developed by Tetra Tech in a coordinated effort with City of San Diego staff, SCCWRP, and staff from the Regional Board as a rapid screening tool to help analysts identify likely causes of biological impairment in a given waterbody and rule out pollutants (e.g., certain pesticides) or other types of stressors (e.g., habitat constraints) that are unlikely to cause impairment at the site of interest. CASTool results indicate which stressor(s) are likely causes of biological impairment using a weight-of-evidence approach and multiple lines of evidence, similar to EPA's Causal Analysis Diagnostic and Decision Information System (CADDIS; USEPA (2017)).

A reach might contain several or no bioassessment sites; if sites are located on the reach, they may have been sampled once or many times. Stressor data used as input to the CASTool are obtained from the SMC data portal and include all the available observed water chemistry data, calculated physical habitat metrics and final Index of Physical Integrity (IPI). Stressor data also include modeled flow metrics (obtained October 2019, personal communication with SCCWRP staff). The CASTool uses stressor data corresponding with response data collected within 10 days of the stressor sample collection for non-modeled data, and all modeled data for sites also having response data. Stressor data available in the SMC portal span the time frame from 2000 to 2018 (Figure 3-4, SCCWRP Staff (2019)). Only likely causes of biological impairment are used as input into the RPPTool for each reach for which CASTool results are available. The user may limit the range of data desired for use in the RPPTool.

Number of stressor samples collected each year
SMC dataset
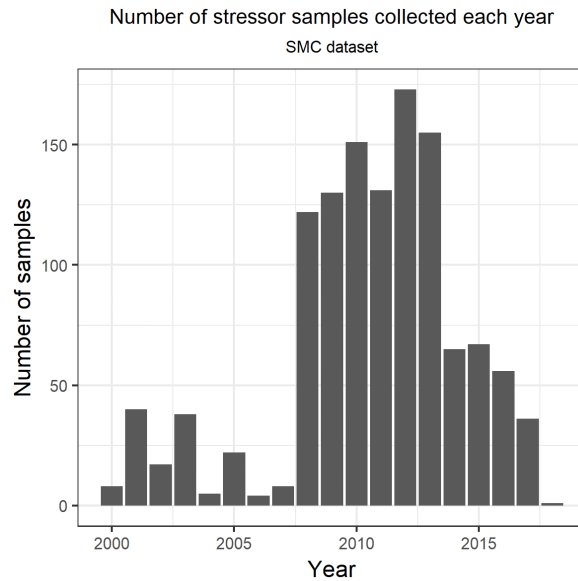
Figure 3-4. Number of stressor samples in the SMC dataset by year between 2000 and 2017.

3.6 Land Cover

The San Diego Association of Governments (SANDAG) has responsibility for regional planning. Every four years, SANDAG updates the regional forecast for planned land use, updating the geographic data each time. The most recent publicly available geographic data for land use forecasts for the year 2050 were obtained from the SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal (SANDAG/SanGIS, 2018b). Similarly, current land use data (2018) were obtained from the same warehouse (SANDAG/SanGIS, 2018a). Current land use information is updated regularly using aerial photography, the County Assessor Master Property Records file, and other ancillary information. The land use information is also reviewed by each of the local jurisdictions and the County of San Diego to ensure its accuracy. This inventory contains more categorical detail and has better positional accuracy than previous county-wide land use inventories.

Land use threats to stream communities often arise from development that results in increased impervious surface and increased urban runoff. To represent these threats, we used the calculated change in combined land use categories representing developed area in each reach catchment between current (2018) and future (2050) land uses. The current land use dataset is also used to identify parks, open space, and low-impact commercial recreation lands that may represent opportunities for successful restoration or protection. These recreational and open space land areas are combined to determine the areal contribution of parks within the catchment.

3.7 Fire Hazard

The incidence and severity of fires has increased dramatically in California in recent years and is predicted to continue as one of the consequences of climate change. In addition to the devastating effects on people and property, large fires can damage vegetation and soils, leading to erosion and other adverse impact on watersheds and their streams. Fire hazard is the likelihood that an area will burn over a 30- to 50-year period without considering modifications such as fuel reduction efforts. Fire risk is the potential damage in the same area resulting from a fire under existing conditions. Fire damage incorporates metrics such as financial impact and housing loss, in addition to damage to natural vegetation, and therefore isn't specifically applicable to biological condition of streams. The RPPTool uses fire hazard as a better representation of the threat to biological stream condition from fires.

California's Department of Forestry and Fire Protection, Fire and Resource Assessment Program (CalFire) has modeled and mapped areas of moderate, high, and very high fire hazard based on vegetation, topography, weather, crown fire potential, ember production and movement, and likelihood. CalFire provides statewide geographic data identifying fire hazard (CalFire, 2007). The RPPTool incorporates these data for San Diego, Kern, Riverside, Orange, Los Angeles, Ventura, and San Bernardino Counties, where they overlap with the SMC region stream reaches. Fire hazard combines the percent of very high, high, and (optionally) moderate fire hazard area in the catchment, which presents a threat to the success of restoration or protection management actions.

3.8 Native Aquatic Species Vulnerability Index

The Native Aquatic Species Vulnerability Index (NASVI) is a normalized average of two metrics, the Rarity Index (which assesses the average rarity of the species in a watershed with small-range species being considered rare) and the count of threatened and endangered species in the watershed (USEPA, 2018). Areas with a higher NASVI provide a greater opportunity for co-benefits if the reach is prioritized for restoration or protection. The data provided were generated for the HUC12 scale. To generate catchment-scale data, the dataset was intersected with the catchments and an area-weighted index average was determined for each catchment. Consequently, the NASVI value for a catchment may over- or under-estimate the NASVI for that reach. However, since native aquatic species are central to the goals of RPP, these data are also included as a positive co-benefit.

3.9 Multiple Species Conservation Program

The Multiple Species Conservation Program (MSCP), established in 1997, is an integral part of regional efforts in San Diego County to protect open space and native species(City of San Diego, 1997). The goal of the MSCP (a 50-year program) is to maintain and enhance biological diversity in the region and maintain viable populations of endangered, threatened, and key sensitive species and their habitats. Not only are endangered and threatened species protected under the MSCP, but residents of the region benefit from the preservation of the natural environment as well. The MSCP is a cooperative effort among the County, the

City of San Diego and other local jurisdictions, along with the U.S. Fish and Wildlife Service and the California Department of Fish and Wildlife (California Department of Fish and Wildlife, 2020). These public partners work with various private landowners, conservation groups, community planning groups, developers and other stakeholders to carry out MSCP objectives.

The MSCP includes the Multi-Habitat Planning Area (MHPA), which is the County's planned habitat preserve and Environmentally Sensitive Areas (ESAs). ESAs include conservation and mitigation lands, but also waters that contain rare, threatened, or endangered species, and Areas of Significant Biological Concern. The geographic data are available for download from the SanGIS/SANDAG GIS Data Portal (SANDAG/SanGIS, 2015). The data are provided in three parts: MSCP_CN for the South County Subregional Plan (2013), MSCP_EAST_DRAFT_CN for the Draft East County Subregional Plan (2009) and MSCP_NORTH_DRAFT_CN for the Draft North County Subregional Plan (2017). Data from each subregion were overlaid on the SMC region catchments to determine the area of protected or preserve land within each catchment. These areas provide an opportunity for co-benefits if a reach within MCSP Plan boundaries is prioritized for restoration or protection.

4.0 RPPTool Methods

The RPPTool is comprised of a Shiny interface (Chang et al., 2019), an R-based web application framework, with the underlying calculations developed in the R programming language R Core Team (2018).

The RPPTool calculates the restoration and protection index scores for each selected reach by averaging the scores from each of the three subindices: Potential, Threats, and Opportunities. Each subindex score is the normalized, weighted average of its component indicator scores, including any user-specified weighting, calculated separately for protection and restoration (specific indicator scoring formulas are shown in the separate subsections below). RPPTool users may apply different weights to each the subindex score, if desired, when calculating the overall index (Equation 1). The subindex scores are presented alongside the overall RPP index scores along with ranks comparable to EPA's RPS Tool. The default weight for each subindex is one (i.e., no weighting).

Equation 1. Calculation of RPP Index incorporating user-weighting for each subindex score.

$$RPP\ Index_{Restore\ OR\ Protect} = \frac{(wt_{potential} \times Sub_{restore\ OR\ protect}) + (wt_{threat} \times Sub_{threat}) + (wt_{opp} \times Sub_{opp})}{wt_{potential} + wt_{threat} + wt_{opp}}$$

where:

wt = user-specified weight for specified subindex, default equals 1

Sub = value of specified subindex (opp = opportunity)

4.1 Potential Subindex

The Potential Subindex score for a reach is a measure of the potential ecological integrity that can be attained by either of two different types of management activities, restoration or protection. The Potential Subindex is made up of four indicators—Biological Condition, Biological Connectivity, Stressors, and Stressor Connectivity—as described below (Figure 4-1).

The Biological Condition Indicator is calculated independently for each type of management activity, therefore, two different Potential Subindex scores are calculated: one for restoration and one for protection. The Potential Subindex equations also differ depending on the availability of stressor data. If no stressor data are available, then the Potential Subindex score is simply the weighted average of the Biological Condition and Biological Connectivity Indicators. If stressor data are available and used in the RPPTool, then the Potential Subindex is the weighted average of the four component indicators: Biological Condition, Biological Connectivity, Stressor, and Stressor Connectivity (Equation 2). The default weight for each indicator is one.

Equation 2. Potential Subindex, calculated separately for restoration and protection potentials.

$$Potential\ subindex_{Restore\ OR\ Protect} = \left(\frac{\sum_{i=1}^{n} weight_i \times value_i}{\sum_{i=1}^{n} weight_i} \right)$$

where:

weight = user-specified weight for indicator "i", default equals 1

value = value of indicator "i"
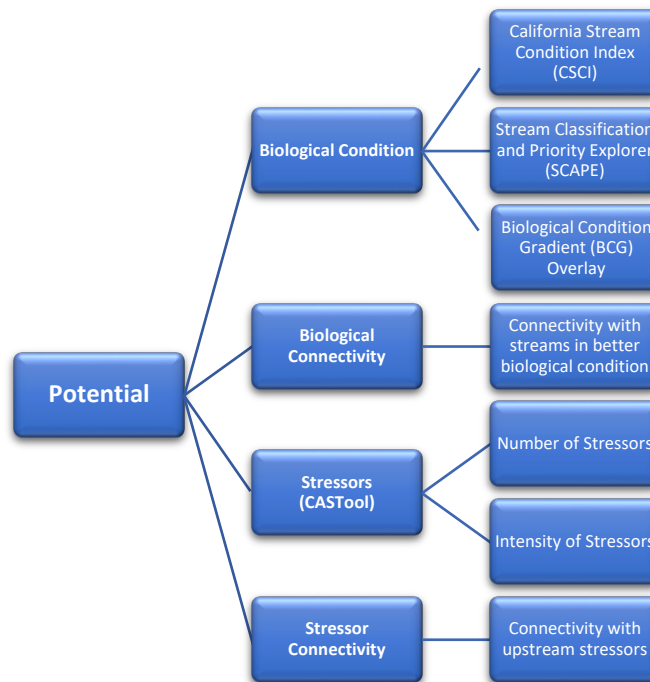
Figure 4-1. Potential Subindex and its component indicators and data sources.

4.1.1 Biological Condition Indicator

The Biological Condition Indicator overlays the California state-wide biological condition gradient (BCG) model on actual or predicted California Stream Condition Index (CSCI) values to determine the biological condition as described by the six-tier BCG from "fully natural" to "extreme changes in community structure and ecosystem function." This indicator incorporates the distribution of predicted CSCI scores for each reach using landscape constraints as modeled in the Stream Classification and Priority Explorer (SCAPE) tool developed by the Southern California Coastal Water Research Project (SCCWRP) for the Stormwater Monitoring Coalition (SMC). These predicted CSCI distributions are compared to observed site CSCI scores. This comparison indicates a site's potential for achieving or maintaining an improved condition relative to predictions. Observed CSCI scores that are relatively low compared to the predicted distribution of CSCI scores for the reach given observed landscape constraints have a greater potential for ecological uplift (i.e., restoration); CSCI scores that are relatively high compared to the predicted distribution of CSCI scores for the reach given observed landscape constraints have a greater value for protection. If no biological

assessment sites occur on a stream reach being evaluated, the Biological Condition Indicator uses only the predicted distribution of CSCI scores and related BCG levels to calculate an indicator score.

4.1.1.1 Reaches with CSCI observations

For reaches with both observed and predicted CSCI scores, two components are calculated. The first component ($CSCI_{protect\ or\ restore}$) compares the observed value to the predicted distribution to determine whether the observed value can increase or decrease to a different third (approximately) of the distribution (Table 4-1). Figure 4-2 illustrates this score. The second component ($BCG_{obs,\ protect\ or\ restore}$) evaluates the maximum possible change that could result in the BCG tier changing to the maximum predicted tier. The direction of change is different for protection or restoration actions (Table 4-1-1).

Table 4-1-1. Biological condition component comparing observed and predicted CSCI scores.

$CSCI_{obs}$ category (of $CSCI_{pred}$ distribution)	Restoration Score	Protection Score
$CSCI_{obs}$ <35th percentile ($CSCI_{qt35}$)	+2	0
$CSCI_{qt35}$ ≤ $CSCI_{obs}$ ≤ $CSCI_{qt65}$	+1	+1
$CSCI_{qt65}$ <$CSCI_{obs}$	0	+2
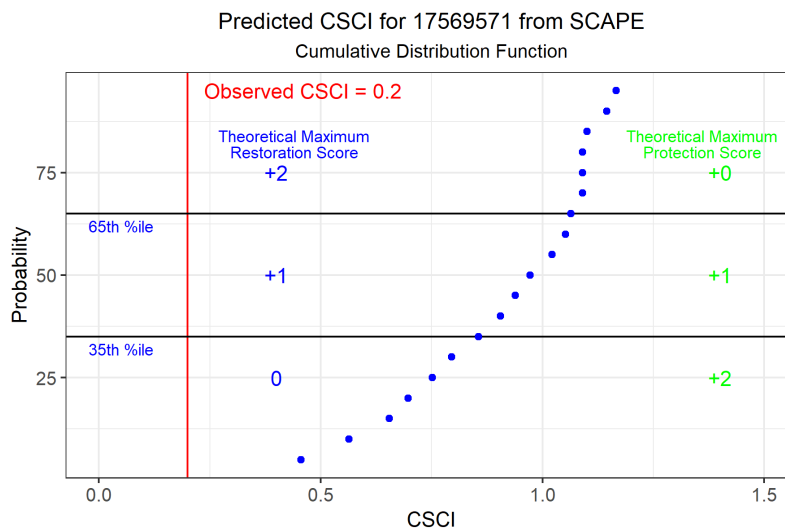
Figure 4-2. Visualization of the biological condition component comparing observed and predicted CSCI values.

The example reach in Figure 4-2 receives a restoration component score of +2, because the observed CSCI score, if restored to its maximum potential, would move from the bottom third of the predicted distribution to the top third. Similarly, the protection component score is 0 because a protection score is applied in the opposite direction. In this example, the reach is already in the bottom third and protection is therefore a lower priority.
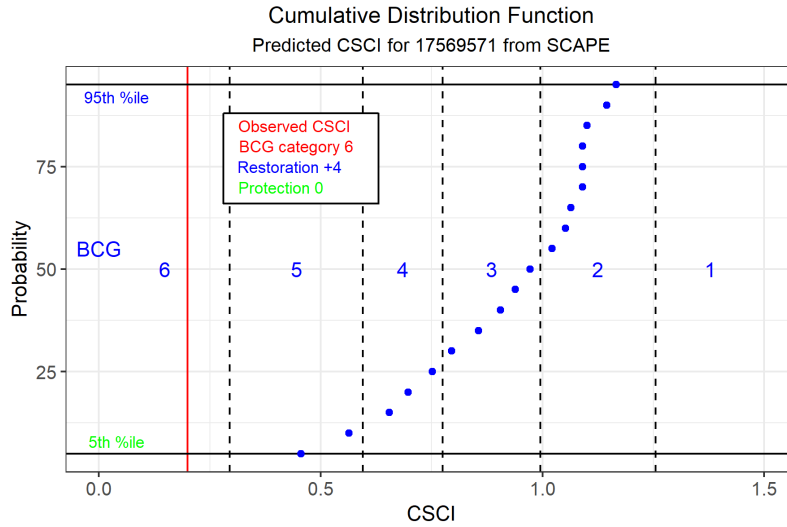
Figure 4-3. Visualization of the biological condition component crediting improvements to BCG tier. The second component relies on the BCG tier that the observed CSCI score falls within and compares it to the BCG tier corresponding with the 95th percentile of the predicted CSCI distribution (for restoration) or the 5th percentile (for protection). The component score is simply the difference between the observed tier and the predicted tier (Figure 4-3). Negative scores are not realistic, so this component is truncated at zero.

4.1.1.2 Reaches without CSCI observations

The Biological Condition Indicator score for reaches without bioassessment sites can still be calculated, but the score relies on the predicted distribution of CSCI scores for the reach. Several reaches do not have predicted CSCI distributions; therefore, biological condition scores for these reaches cannot be calculated. Using the predicted CSCI distribution, the biological condition score for restoration is the difference between the BCG tier corresponding with the median predicted CSCI and the BCG tier corresponding with the 95th percentile (**Equation 3**). The biological condition score for protection is the difference between the BCG tier corresponding with the 5th percentile predicted CSCI and the median (**Equation 4**).

Equation 3. Biological condition score (restoration) for reaches without observations.

$$BCG_{pred,restore} = BCG_{qt50} - BCG_{qt95}$$

where:

BCG_{qtx} = BCG tier corresponding with the xth percentile

Equation 4. Biological condition score (protection) for reaches without observations.

$$BCG_{pred,protect} = BCG_{qt05} - BCG_{qt50}$$

where:

BCG_{qtx} = BCG tier corresponding with the xth percentile

4.1.1.3 BCG Bonus (optional)

Some reaches, like the one in the example above (Figure 4-2 and 4-3), may have an observed BCG tier less than the BCG tier corresponding with the 5th percentile of the predicted CSCI distribution (BCG$_{qt05}$). Others may have an observed BCG tier greater than the BCG tier corresponding with the 95th percentile (BCG$_{qt95}$). In the former case, if the BCG$_{qt05}$ is tier four (4) or greater, then the significant improvement to a tier in which "ecosystem functions are largely maintained" might have greater benefit than a different reach where the observed BCG tier is within the distribution of predicted values. Conversely, for the latter case, if the BCG$_{qt95}$ is tier two (2) or better (structure and function are similar to the natural community) and the observed BCG tier is better than the BCG$_{qt95}$, then protecting this reach from degradation might have great benefit. The user can decide whether to attach a "bonus" score (+1) in these cases. If this option is selected, it applies to both the Restoration and Protection Indicator components.

4.1.1.4 Headwater Bonus (optional)

As an option, the user may decide to attach a "bonus" point (+1) to reaches that are headwaters. Headwater streams in populated regions such as the SMC region are frequently at risk from various anthropogenic activities and resulting stressors. Thus, a user may want to know that a reach is a headwater stream and apply a bonus to its indicator score. The "StartFlag" field drawn from the NHDPlus flowline table equals one (1) if the reach is a headwater and zero (0) if the reach is not a headwater. If the user opts to allow the headwater bonus, the +1 value is added based on the value of this field for each reach. If this option is selected, it applies to both the Restoration and Protection Indicator components.

4.1.1.5 Final Biological Condition Indicator Scores

The final Biological Condition Indicator scores (one for Restoration and one for Protection) are calculated as the arithmetic average of the component scores, for reaches with observed biological condition (measured CSCI value) and predicted biological condition (SCAPE data only) separately, as shown by **Equation 5** and **Equation 6**. The same equation is used for restoration and protection. The difference in the equations depends only on the availability of observational data.

Equation 5. Biological Condition Indicator score for reaches with bioassessment sample data.

$$Indicator_{BioCond,obs} = \frac{CSCI_{obs} + BCG_{obs} + Bonus_{HW} + Bonus_{BCG}}{BCG_{max} + Bonus_{HW,max} + Bonus_{BCG,max}}$$

where:

CSCI$_{obs}$ = CSCI score as described in Table 4-1-1 and *Figure*

BCG$_{obs}$ = BCG score as described in Figure

Bonus$_{HW}$ = Headwater bonus value (0 or 1), if used

Bonus$_{BCG}$ = BCG bonus value (0 or 1), if used

BCG$_{max}$ = Maximum BCG (6)

Bonus$_{HW,max}$ = Maximum headwater bonus (1 if used, 0 if not)

Bonus$_{BCG}$ = Maximum BCG bonus (1 if used, 0 if not)

Equation 6. Biological Condition Indicator score for reaches without bioassessment sample data.

$$Indicator_{BioCond,pred} = \frac{BCG_{pred} + Bonus_{HW} + Bonus_{BCG}}{BCG_{max} + Bonus_{HW,max} + Bonus_{BCG,max}}$$

where:

BCG_{pred} = BCG tier as described in

Equation 2 or Equation 3

$Bonus_{HW}$ = Headwater bonus value (0 or 1), if used

$Bonus_{BCG}$ = BCG bonus value (0 or 1), if used

BCG_{max} = Maximum BCG (6)

$Bonus_{HW,max}$ = Maximum headwater bonus (1 if used, 0 if not)

$Bonus_{BCG}$ = Maximum BCG bonus (1 if used, 0 if not)

4.1.2 Biological Connectivity Indicator

The Biological Connectivity Indicator represents the hydrologic connectivity of the target reach to other streams in better biological condition, determined using the mapped stream reaches. Connectivity with better streams reflects the proximity of the reach to source populations that may facilitate biological uplift from restoration or maintain biological condition if protected, therefore raising the biological condition score for the Potential Subindex. The default stream length used to identify contributing source populations is 5 km but the user may modify the length used to designate connected reaches. Connectivity with streams in better biological condition is measured both upstream and (optionally) downstream from either the top (for upstream) or bottom (for downstream) of the target reach. The target reach is excluded from the individual connected reach length or aggregate connected reach length.

To calculate the Biological Connectivity Indicator, observed data are given priority over predicted data in the order shown in Table 4-2. Only connected reaches with better biological condition are included in the Biological Connectivity Indicator because reaches of similar or worse biological condition are unlikely to contribute to biological uplift at the site of interest. Figure 4-4 presents a map of the same reach depicted in the Biological Condition Indicator graphics. Predicted (median) and observed BCG tiers are shown using different colors. Connected "better than" BCG tiers are outlined and are used in the numerator to calculate the Biological Condition Indicator score, as shown in **Equation 7**. The aggregate length in the denominator includes all connected reaches.

Table 4-1-2. Priority of observational versus predicted data for the Biological Connectivity Indicator score.

Priority	Target Reach	Connected Reaches
1	Observed CSCI/BCG	Observed CSCI/BCG
2	Observed CSCI/BCG	Predicted $CSCI_{qt50}$/BCG_{qt50}
3	Predicted $CSCI_{qt50}$/BCG_{qt50}	Observed CSCI/BCG
4	Predicted $CSCI_{qt50}$/BCG_{qt50}	Predicted $CSCI_{qt50}$/BCG_{qt50}
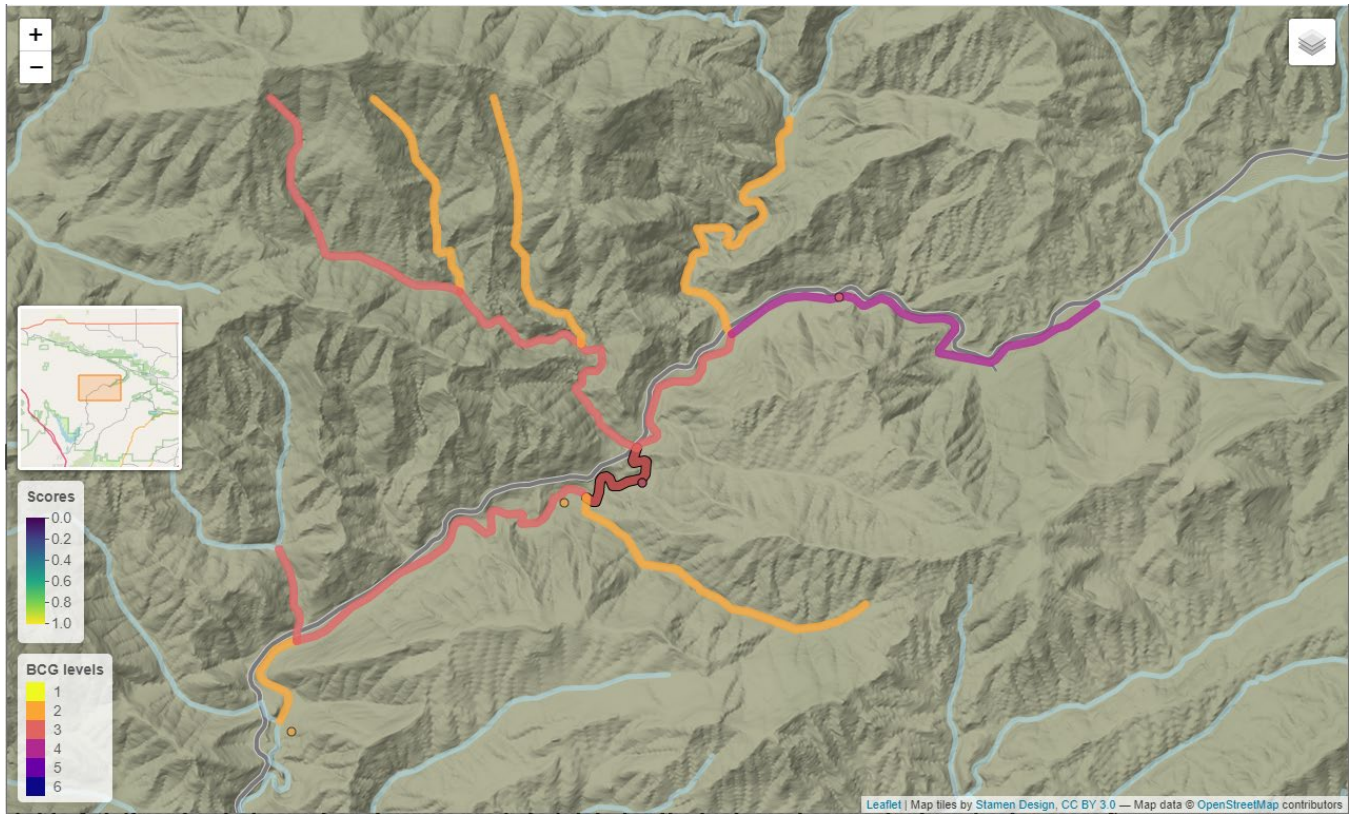
Figure 4-4. Biological connectivity map for reach 17569571. BCG_{qt50} tiers for reaches in black; observed BCG scores for tiers on the reaches in orange. Dots with arrows represent biological sampling points.

Equation 7. Biological Connectivity Indicator score calculation. Only "better than" biological quality reaches contribute to the numerator.

$$Indicator_{BioCxn} = \sum_{i=1}^{n} \left(\frac{\left(BCG_{target} - BCG_i\right) \times length_i}{max(BCG_{delta}) \times length_{total}} \right)$$

where:

n = total number of connected reaches

BCG_{target} = BCG tier for the target reach (observed or predicted)

BCG_i = BCG tier for the connected reach (observed or predicted)

$length_i$ = Length of the connected reach

$max(BCG_{delta})$ = Maximum difference in BCG scores (5)

$length_{total}$ = Aggregate length of all connected reaches

4.1.3 Stressor Indicator

The Stressor Indicator uses likely candidate causes affecting biological condition at bioassessment sites that are located on the reach, if any. These likely candidate causes (stressors) are identified using the companion screening causal assessment tool (CASTool). Stressors are defined as chemical, flow, or habitat measures that are identified as falling outside the interquartile range of comparator samples and having supporting weight of evidence as a cause of benthic macroinvertebrate community impairment. The number of stressors present and the intensity of each stressor at each site are combined into standard scores that can depress the biological condition score in the Restoration and Protection Potential Subindex calculations. The user can choose to weight each stressor found as 1 (default), 0 (not considered), or 2 (twice as important).

The RPPTool will not calculate the Stressor Indicator under the following conditions: if the CASTool is not run for any sites located on the selected reach prior to running the RPPTool; if appropriate stressor and response data are not available for the reach; or no stressors are identified by the CASTool. In these cases, the RPPTool will calculate other indicators that do not rely on stressor information. Lack of stressor data for a reach is noted in the RPPTool output.

The calculations for the Stressor Indicator require several steps, similar to the approach used in EPA's RPS Tool. First, the observed stressor values are scaled suing the full range of values observed in the dataset, resulting in values ranging from zero to one for each stressor independently of others. Second, the user-specified weight is applied (Table 4-3). Lastly, the scaled, weighted sum of stressors is calculated. This result is subtracted from one to obtain the final Stressor Indicator score (**Equation 8**), to account for the observation that more stressors or greater intensity of stressors will result in poorer biological condition.

Table 4-1-3. Example user weighting scheme for selected stressors.

Stressor Group	Label	Weight
Habitat	Evenness of flow habitat types	1
	Index of physical habitat integrity	2
	Riparian cover (sum of three layers)	1
Modeled flow metrics	Wet-season maximum mean monthly streamflow (m3/s)	1
	Average 99t percentile of daily streamflow (m3/s)	1
	Dry-season Richards-Baker Index (flashiness)	1
Nutrients	Particulate chlorophyll a (mg/m2)	1
	Dissolved nitrite as N (mg/L)	1
	Total phosphorus as P (mg/L)	2
	Total nitrogen (mg/L)	2
Water Quality	Dissolved alkalinity as calcium carbonate (mg/L)	0
	Field-measured dissolved oxygen (mg/L)	1
	Field-measured specific conductivity (µS/cm)	1
	Total dissolved solids (mg/L)	1
	Field-measured water temperature (degrees C)	1

Equation 8. Stressor Indicator score calculation.

$$Indicator_{Stressor} = 1 - \sum_{i=1}^{n} \left(\frac{weight_i \times value_i}{\sum_{i=1}^{n} weight_i} \right)$$

where:

n = number of stressors at the target reach

weight = user-specified weight for stressor "i", default equals 1

value = value of stressor "i" scaled from 0 to 1

4.1.4 Stressor Connectivity Indicator

The Stressor Connectivity Indicator represents the hydrologic connectivity of the reach to other streams with the same stressors. Connectivity with streams with similar stressors reflects the proximity of the reach to upstream stressors that may contribute to continued degradation and, therefore, can depress the depress the biological condition. The weights selected by the user for each stressor are the same as the weights applied for the stressor indicator. The stream length used to identify upstream stressors is identical to the connectivity distance selected for biological connectivity (default = 5 km), which also can be modified by the user but is always equal for the Biological Connectivity Indicator and the Stressor Connectivity Indicator. Like the Stressor Indicator, the RPPTool will not calculate the Stressor Connectivity Indicator if results from the CASTool are unavailable for any of the reasons previously stated.

Stressors used for the Stressor Connectivity Indicator are the stressors that are identified as likely causes of biological impairment at the target reach. Although other stressors may be likely causes of impairment at the connected reaches, the RPPTool addresses only the contributors to stressors documented and believed to be responsible for impairment at the target reach. The indicator calculation creates a length-weighted average of the stressor indicator for each upstream reach, as shown in **Equation 9**. As for the Stressor

Indicator, because the number of stressors and stressor intensity may adversely affect biological condition, the final indicator score equals the length-weighted sum of scaled stressors subtracted from one.

Equation 9. Stressor Connectivity Indicator calculation for stressors identified at the target reach.

$$Indicator_{StressCxn} = 1 - \sum_{i,j=1}^{n,m} \left(\frac{weight_i \times value_{i,j}}{\sum_{i=1}^{n} weight_i} \right) \times \left(\frac{length_j}{\sum_{j=1}^{m} length_j} \right)$$

where:

n = total number of stressors at the target reach

m = total number of connected reaches

weight = user-specified weight for stressor "i", default equals 1

value = value of stressor "i" scaled from 0 to 1 at the connected reach j

length = length of the connected reach j

4.2 Threats Subindex

The Threats Subindex for a reach is a measure of current and future conditions that can adversely affect the success of restoration or protection measures implemented to improve or maintain ecological integrity. This subindex is made up of two threat types—Land Use Change and Fire Hazard—as described below (Figure 4-5). RPPTool users may weight each indicator in the subindex, with the final subindex calculated as the weighted average of the individual indicators, this result is subtracted from one to obtain the final Threats Subindex score, accounting for the observation that increasing threats result in poorer biological condition (Equation 10). The default weight for each indicator is one.

Equation 10. Threats Subindex calculation incorporating user weighting for each indicator.

$$Threat\ subindex = 1 - \left(\frac{\sum_{i=1}^{n} weight_i \times value_i}{\sum_{i=1}^{n} weight_i} \right)$$

where:

weight = user-specified weight for indicator "i", default equals 1

value = value of indicator "i"

Figure 4-5. Threats Subindex with indicators and data sources.

4.2.1 Land Use Change Indicator

Land use is the stressor most closely associated with ecological integrity across ecosystems, including streams. Human land uses at multiple scales including reach riparian area, reach catchment, and broader watershed have all been shown to adversely affect biological condition. Therefore, land use change is likely the greatest future threat to biological condition of streams that are being considered for restoration or protection. While land use change predictions are uncertain, detailed planned land use geographic data layers for 2050 covering the San Diego region were compared to current land uses in the same region (Source: (SANDAG/SanGIS)). Projected increases in developed land uses in the reach catchment represents a negative effect on biological condition. Table 4-4 presents the land use categories that combine as developed lands.

Table 4-1-4. Developed land categories.

General Category	Specific Planned Land Use Code	Specific Current Land Use Code
Airports	Airstrip	Airstrip
	Commercial airport	Commercial airport
	General aviation airport	General aviation airport
	Military airport	Military airport
Commercial use	Arterial Commercial	Arterial Commercial
	Automobile Dealership	Automobile Dealership
		Commercial under Construction
	Community Shopping Center	Community Shopping Center
	Neighborhood Shopping Center	Neighborhood Shopping Center
	Other Retail Trade and Strip Commercial	Other Retail Trade and Strip Commercial
	Regional Shopping Center	Regional Shopping Center
	Service Station	Service Station
	Specialty Commercial	Specialty Commercial
	Wholesale Trade	Wholesale Trade
Commercial Recreation	Casino	Casino
	Convention Center	Convention Center
	Golf Course	Golf Course
	Golf Course Clubhouse	Golf Course Clubhouse

General Category	Specific Planned Land Use Code	Specific Current Land Use Code
	Olympic Training Center	Olympic Training Center
	Other Recreation - High	Other Recreation - High
	Racetrack	Racetrack
	Stadium/Arena	Stadium/Arena
	Tourist Attraction	Tourist Attraction
Extractive Industry	Extractive Industry	Extractive Industry
Group Quarters	Dormitory	Dormitory
	Jail/Prison	Jail/Prison
	Military Barracks	Military Barracks
	Monastery	Monastery
	Other Group Quarters Facility	Other Group Quarters Facility
Heavy Industry	Heavy Industry	Heavy Industry
Hospitals	Hospital - General	Hospital - General
	Other Health Care	Other Health Care
	UCSD/VA Hospital/Balboa Hospital	UCSD/VA Hospital/Balboa Hospital
Hotel/Motel/Resort	Hotel/Motel (High-Rise)	Hotel/Motel (High-Rise)
		Hotel/Motel (Low-Rise)
	Resort	Resort
Junkyard/Dump/Landfill	Junkyard/Dump/Landfill	Junkyard/Dump/Landfill
Light Industry	Industrial Park	Industrial Park
		Industrial under Construction
	Light Industry - General	Light Industry - General
	Public Storage	Public Storage
	Warehousing	Warehousing
Military	Military Training	Military Training
	Military Use	Military Use
	Weapons Facility	Weapons Facility
Mobile Home Park	Mobile Home Park	Mobile Home Park
Multi-Family Residential	Multi-Family Residential	Multi-Family Residential
	Multi-Family Residential Without Units	Multi-Family Residential Without Units
	Single Room Occupancy Units (SRO's)	Single Room Occupancy Units (SRO's)
Office	Government Office/Civic Center	Government Office/Civic Center
	Office (High-Rise)	Office (High-Rise)
	Office (Low-Rise)	Office (Low-Rise)
		Office under Construction
Other Transportation	Communications and Utilities	Communications and Utilities
	Freeway	Freeway
		Freeway under Construction
	Marine Terminal	Marine Terminal
	Other Transportation	Other Transportation
	Park and Ride Lot	Park and Ride Lot
	Parking Lot – Structure	Parking Lot - Structure
	Parking Lot – Surface	Parking Lot - Surface
Other Transportation (cont.)	Rail Station/Transit Center	Rail Station/Transit Center
	Railroad Right of Way	Railroad Right of Way
	Road Right of Way	Road Right of Way

General Category	Specific Planned Land Use Code	Specific Current Land Use Code
		Road under Construction
Public Services	Fire/Police Station	Fire/Police Station
	Library	Library
	Mission	Mission
	Other Public Services	Other Public Services
	Post Office	Post Office
	Religious Facility	Religious Facility
Residential/Commercial mixed use	Mixed Use	Mixed Use
Schools	Elementary School	Elementary School
	Junior College	Junior College
	Junior High School or Middle School	Junior High School or Middle School
	Other School	Other School
	Other University or College	Other University or College
	School District Office	School District Office
		School under Construction
	SDSU/CSU San Marcos/UCSD	SDSU/CSU San Marcos/UCSD
	Senior High School	Senior High School
Single Family Residential		Residential under Construction
	Single Family Detached	Single Family Detached
	Single Family Multiple-Units	Single Family Multiple-Units
	Single Family Residential Without Units	Single Family Residential Without Units
Spaced Rural Residential	Spaced Rural Residential	Spaced Rural Residential
		Spaced Rural Residential Without Units

Developed land area in the reach catchment was determined as a fraction of the total catchment area for current land use (2018) and planned land use (2050) separately. The summed developed land area (current, 2018) was subtracted from the summed developed land area (planned, 2050) to determine the change in land use (Equation 11). Since all values are fractions of the total catchment area, the final values range from zero (0) to one (1).

Equation 11. Change in land use indicator calculation for each reach catchment.

$$Indicator_{ChangeDevLU} = \left(\sum_{i=1}^{n} Fraction\ Developed\ LU\ 2050_{category} \right) - \left(\sum_{i=1}^{n} Fraction\ Developed\ LU\ 2018_{category} \right)$$

where:

n = total number of developed land use categories

Fraction Developed LU = fraction of developed land use in the reach catchment

4.2.2 Fire Hazard Indicator

Fire hazard areas are rated as very high likelihood, high likelihood, or moderate likelihood (Source: (CalFire)). The areas within each catchment that are rated in these categories was determined and the fraction of total catchment area in each category was calculated. Users can select whether the Fire Hazard Indicator should include moderate likelihood of fire. The Fire Hazard Indicator is calculated as the sum of catchment area meeting the hazard criteria (Equation 12).

Equation 12. Fire hazard indicator calculation for each reach catchment.

$$Indicator_{FireHazard} = \sum_{i=1}^{n} \frac{Fire\ hazard\ area_i}{Total\ catchment\ area}$$

where:

Fire hazard area = fraction of fire hazard of the specified likelihood (i) within the catchment

4.3 Opportunities Subindex

The Opportunities Subindex score for a reach is a measure of the co-benefits that may be attained, in addition to ecological integrity, by choosing to restore or protect. Co-benefits may occur in a variety of forms, including (1) conservation benefits to biodiversity that may be realized with restoration or protection of a reach, (2) recreational opportunities by improving or protecting streams near parks, and (3) potential synergies with other on-going efforts where there is stakeholder interest, such as Watershed Management Plans. The conservation, recreational, and other co-benefits that can be obtained by co-locating restoration or protection are included in the Opportunities Subindex as three opportunity types comprised of four separate indicators as depicted below—Species Conservation (Multiple Species Conservation Program Area and the Native Aquatic Species Vulnerability Indicators), Recreational Co-benefits (Parks Indicator), and User-defined Co-benefits Indicator (Figure 4-6). RPPTool users may weight each indicator in the subindex, with the final subindex calculated as the weighted average of the individual indicators (Equation 13). The default weight for each indicator is one.
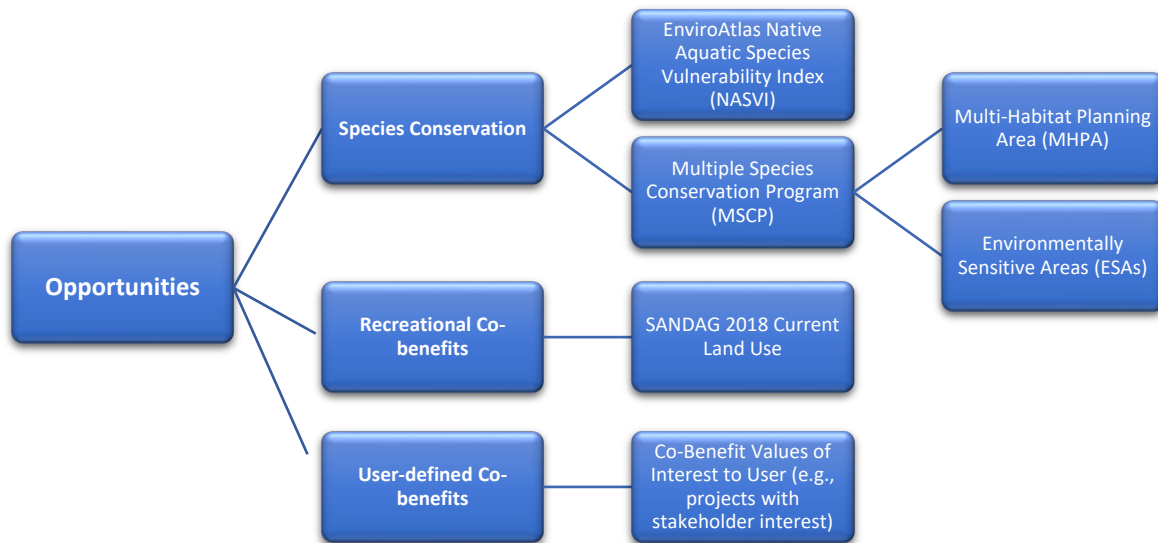
Equation 13. Opportunities Subindex calculation incorporating user-weighting for each indicator.

$$Opportunity\ subindex = \left(\frac{\sum_{i=1}^{n} weight_i \times value_i}{\sum_{i=1}^{n} weight_i} \right)$$

where:

weight = user-specified weight for indicator "i", default equals 1

value = value of indicator "i"

Figure 4-6. Opportunities Subindex with indicators and data sources.

4.3.1 Species Conservation Indicators

The first of two indicators representing species conservation is EnviroAtlas' Native Aquatic Species Vulnerability Index (NASVI). This indicator is a normalized average of the average rarity of a species based on their ranges and the count of threatened and endangered species in the watershed (USEPA, 2019). The NASVI was developed for EnviroAtlas at the HUC12 watershed scale (12-digit Hydrologic Unit Code) and extrapolated to represent an approximate index value for the reach catchment. Equation 14 depicts the calculation of this indicator.

Equation 14. Calculation of the Native Aquatic Species Vulnerability Index Indicator.

$$Indicator_{NASVI} = \frac{\sum_{i=1}^{n}(value_i \times area_i)}{area_{catchment}}$$

where:

$value_i$ = NASVI value "i" defined for a specific area of a catchment

$area_i$ = area within a catchment having NASVI value "i"

$area_{catchment}$ = total area of the catchment

The second indicator representing species conservation, San Diego County's MSCP, is the most comprehensive designation of species conservation areas that might benefit from restoration or protection of their streams (Source: The City of San Diego (1998). This indicator differs from the NASVI in two ways: the MSCP represents protected area selected to be protective for species conservation as opposed to

observations and models of species habitats requiring conservation; and the MSCP area covers only a portion of the SMC region, whereas the NASVI covers the entire continental United States.

The MSCP spatial extent and program plan are divided into three separate regions, with separate identified categories contributing to the MSCP overall area. Table 4-5 presents the categories identified within each region, a brief description of them, and an indication of whether they were included in the final sum of "MSCP area." Equation 15 presents the indicator calculation for each category marked as included.

Table 4-1-5. MSCP categories by region.

Region	Category	Included in indicator?
MSCP North	Gregory Canyon landfill	No
	Other lands	No
	Outside pre-approved mitigation areas (PAMA)	No
	State and federal pre-approved mitigation area (PAMA)	Yes
	Preserve Areas	Yes
	Special Districts	No
	Take Authorized	No
	Tribal Lands in Fee	No
	Tribal Lands in Trust	No
	US Forest Service	Yes
MSCP East	Agriculture or natural upland within focused conservation areas (FCA)	Yes
	Agriculture or natural upland outside focused conservation areas (FCA)	Yes
	Baseline Preserve	Yes
	Developed Lands	No
	No description	No
	Riparian/wetland habitat and transition zone within focused conservation areas (FCA)	Yes
	Riparian/wetland habitat and transition zone outside of focused conservation areas (FCA)	Yes
	RMS 1 - Highest level of ecological protection	Yes
	RMS 2 - Land managed with ecological protection	Yes
	RMS 3 - Land managed as open space	Yes
	RMS 4 - Other public/semi-public lands	Yes
	Tribal lands	No
MSCP South	Conserved subject to agreement with wildlife agencies	Yes
	Hardline preserve	Yes
	Major amendment area	No
	Minor amendment area subject to special considerations	No
	Minor amendment area	No
	State and federal pre-approved mitigation area (PAMA)	Yes
	Santa Fe Valley sensitive biological habitat 'D' designator areas	Yes

Region	Category	Included in indicator?
	Golf course related development allowed in Santa Fe Valley open space II areas	Yes
	Take authorized area	No
	Unincorporated land in Metro-Lakeside-Jamul Segment	No

Equation 15. Calculation of the MSCP Indicator.

$$Indicator_{MSCP} = \sum_{i=1}^{n} \frac{MSCP\ area_i}{Total\ catchment\ area}$$

where:

Land use area = fraction of land use area designated as MSCP co-benefit (i) within the reach catchment

4.3.2 Recreational Co-benefits Indicator

Stream restoration and protection can bring significant co-benefits to human activities, such as active recreation (biking, hiking, and/or equestrian uses), passive recreation, bird watching, scientific research, and nature walks. San Diego County's geographic data for current land use describes multiple categories of parks and low-impact recreation areas where co-benefits are most likely to occur. These categories include relatively natural areas such as beaches, landscaped greenbelt areas, open space, campgrounds, wildlife and nature preserves, and undevelopable natural areas. They also include more carefully managed and sometimes constructed areas such as neighborhood parks and recreation centers (Table 4-6).

Since each land use category is calculated as a fraction of the total catchment area, the contributing categories within each catchment are summed for a total fraction of the catchment that contributes to the recreational co-benefit indicator (Equation 16).

Table 4-1-6. Land uses contributing to the Recreational Co-benefits Indicator.

Current Land Use Categories (Parks, Commercial recreation, or Vacant/undevelopable land)	
Commercial recreation	Marina
	Other Recreation - Low
Park	Beach - Active
	Beach - Passive
	Landscape Open Space
	Open Space Park or Preserve
	Park - Active
	Residential Recreation
	Undevelopable Natural Area
Vacant/Undeveloped Land	Vacant and Undeveloped Land

Equation 16. Recreational Co-benefits Indicator calculation for each reach catchment.

$$Indicator_{Recreation} = \sum_{i=1}^{n} \frac{Land\ use\ area_i}{Total\ catchment\ area}$$

where:

Land use area = fraction of land use area designated for recreational co-benefit (i) within the reach catchment

4.3.3 User-defined Co-benefits Indicator

Many other forms of co-benefits may arise through synergies with other watershed activities that are not available for inclusion in the RPPTool. Therefore, this indicator is user defined; i.e., the user can add one additional indicator with a score of one (1) and user weights ranging from one (1) to three (3), corresponding with the categorical assignments of priority low, medium, high, to capture co-benefits they want to include. Examples are projects that are already designated in existing Watershed Management Plans, reaches where there is significant stakeholder interest, and projects that would further other City goals, such as environmental, social, and economic equity. This indicator value is included in the weighted average calculation of the final RPP index scores.

4.4 Limitations

The RPPTool is designed as a screening tool and therefore the user should be aware of certain limitations of the tool. Table 4-7 identifies these limitations by subindex and indicator. The results that are presented inherently include these limitations, requiring care in interpreting the results. Close observation of the environmental setting of the target reach, along with professional judgment, should allow a critical evaluation of the priorities identified by the tool.

Table 4-1-7. Limitations inherent in calculating subindices.

Subindex	Indicator	Limitation
Potential (restoration OR protection)	Biological Condition	Requires predicted biological condition data from SCAPE; 23% of reaches in the SMC region lack predictions.
	Biological Connectivity	Requires predicted biological condition data from SCAPE; only includes reaches that exist in the NHDPlus Value Added Attributes table.
	Stressors	Requires CAST-identified likely causes of existing or potential impairment for the target reach; limited to available data within the time frame selected.
	Stressor Connectivity	Requires CAST-identified likely causes of existing or potential impairment for the target reach; limited to available data within the time frame selected; only includes reaches that exist in the NHDPlus Value Added Attributes table.
Threat	Land Use Change	Current land use is based on 2018 data; projected future land use is based on models anticipating land use in 2050. The error in either of these datasets has not been quantified and is not static in time. Land use data used for this indicator are only available for the County of San Diego. This indicator will not be evaluated for areas outside the county.
	Fire Hazard	CalFire maps three hazard zones in state responsibility areas (very high, high, and moderate hazard). In local responsibility areas only high and very high hazard zones are mapped. If a user selects the option to include moderate

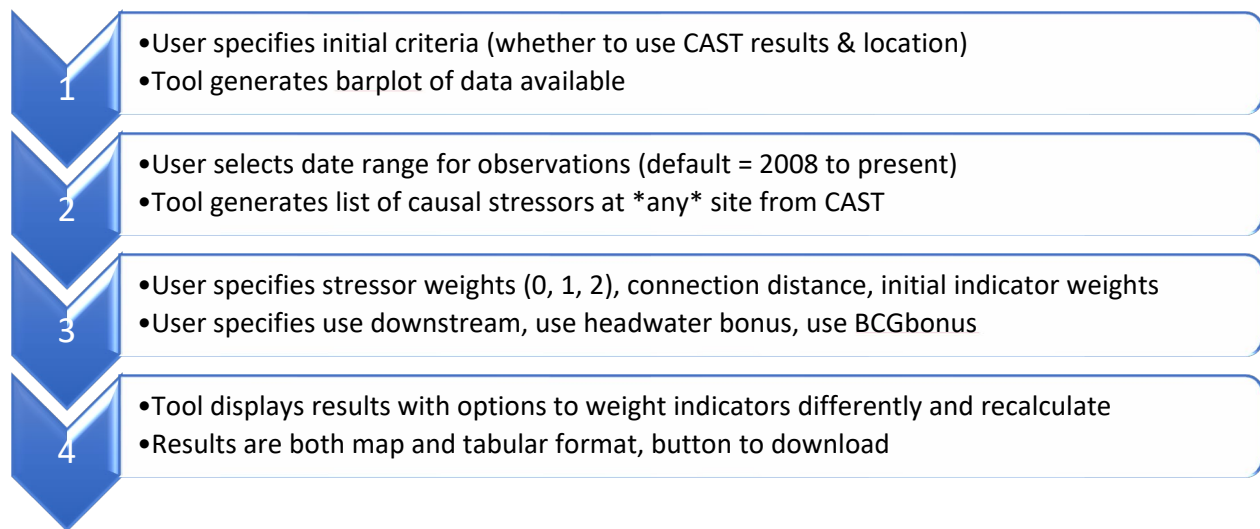| Subindex	Indicator	Limitation
		hazard zones, the Fire Hazard threat indicator may be underrepresented for streams in local responsibility areas. Hazard is the likelihood that an area will burn. It is not risk. Risk includes the potential damage (e.g., to structures) that fire can do under existing conditions.
Opportunity	NASVI	The Native Aquatic Species Vulnerability Index (NASVI) was developed for use at the HUC-12 level. In extrapolating these data to the catchment level, the value published for the HUC-12 was applied equally to each catchment comprising that HUC-12. A given catchment index value may be either under- or overestimated depending on the evenness of distribution of the index across the HUC-12.
	MSCP	Some of the categories defined in the MSCP program may allow limited uses that result in impacts to species intended to be conserved. The amount or location of these hypothetical impacts are not provided in a manner that can be included in an indicator calculation. Therefore, this indicator may over- or underestimate the associated co-benefit.
	Recreational co-benefits	Current land use is based on 2018 data. The error in this dataset has not been quantified and is not static in time (i.e., "current" in 2018 may not be the same as "current" in 2020, at the time of publication of this report). Moreover, land use data for this indicator are only available for the County of San Diego and include a variety of recreational co-benefits. Not all users will agree with the recreational categories included, which may result in this indicator being over- or undervalued.
	User-defined co-benefits	Since user-defined co-benefits are unknown, the limitations must be determined by the user.

5.0 User Experience

The RPPTool is available to users in two separate but compatible formats. The first, a more user-friendly format, is through a web-application framework built using Shiny. The second, suitable for advanced users, requires obtaining and using the R package directly. Table 5-1 compares the major differences between each version.

Table 5-1. Comparison between web-application and R package versions of the RPPTool

Shiny web-based tool	R package
Intuitive web-based application with tabbed navigation	Flexible command-line interface requiring an understanding of R package use
Runs one reach at a time	Runs multiple reaches
Calculates connectivity only for selected (target) reach (all other indicators are run for all reaches; stressor data are included based on availability)	Calculates connectivity for multiple reaches if specified in input file
CASTool results must be uploaded for use in the RPPTool	CASTool results file directory must be identified for use
RPPTool results available as an interactive map and table	RPPTool results available as static maps and a text file

5.1 Process flow

Regardless of the version selected, the process flow is similar (Figure 5-1). For the web-based version, the user must select a specific reach using either the map or the drop-down list prior to following the steps in Figure 5-1. The advanced user will need to specify an Excel file containing the list of reaches to evaluate.

1
- User specifies initial criteria (whether to use CAST results & location)
- Tool generates barplot of data available

2
- User selects date range for observations (default = 2008 to present)
- Tool generates list of causal stressors at *any* site from CAST

3
- User specifies stressor weights (0, 1, 2), connection distance, initial indicator weights
- User specifies use downstream, use headwater bonus, use BCGbonus

4
- Tool displays results with options to weight indicators differently and recalculate
- Results are both map and tabular format, button to download

Figure 5-1. Generalized process flow for both versions of the RPPTool

In both cases, after completing the user input through Step 3 in Figure 5-1, the RPPTool will calculate the following indicators for all reaches for which data are available:

- Potential Subindices (restoration and protection)
    - Biological condition

- - o Stressor

- Threat Subindex

    - o Land use change (increase in developed land)

    - o Fire hazard

- Opportunity Subindex

    - o Native Aquatic Species Vulnerability Index

    - o Multi-Species Conservation Program

    - o Recreational co-benefits

    - o User-defined co-benefits

Connectivity indicators will only be calculated for the reaches selected for analysis, regardless of the interface option used.

The final output contains indicator values that can be calculated for all reaches in the SMC region regardless of data availability; stressor data for all reaches containing sites that have CASTool results and identified stressors; and connectivity indicators for reaches selected for analysis. Subindices are calculated using the available data for each reach in the region, and RPP index scores are generated, both for restoration and protection. An output table (in the web-application interface) or tab-delimited text file that can be opened in Excel present the results.

In the web-based application, user-defined indicator and subindex weights can be altered in the interface and the results recalculated. Since this step does not affect the connectivity calculations that were already executed, the table can be refreshed quickly. In the R package, the function "updateAllScores" can be run with new weights as input to the function.

5.2 Web-application Interface

Although the RPP tool was developed as an R package, using the package requires at least a basic level of proficiency in R. To reach a broader audience, a more user-friendly interface was also required, namely, the shiny interface. Shiny allows the RPPTool to be developed as an app that can be hosted on a website or that the user can run on their desktop. To run the app on the host website, the user only needs to enter the URL into their web browser to access the tool. All user interaction with the code is handled in the web browser by point and click.

The start screen includes a disclaimer from the City (Figure 5-2) with a navigation tool bar that provides entry to the tool and to view the results. Because the user may not know the reach identifier used in the tool, the Shiny app provides two maps to identify the targeted stream reach. The stations map (Figure 5-3) allows the user to select a station ID or to find a stream on a map, similar to Google Earth. Once the targeted stream is located, the user clicks on the reach and a pop-up box displays the name of the stream and the COMID (the stream reach ID). A second map works similarly, but is based on finding locations by

stream reach ID (Figure 5-4) for those cases where the user knows the ID but not where the stream is located. After obtaining the COMID, the user can navigate to the "RPP-Calc" tab to enter the ID and operate the tool (Figure 5-5). The user can run the tool with the base inputs and default criteria, or access other tabs to modify selected criteria before clicking the "Run RPP" button. The RPPTool code (comparable to the package) is executed in the background and the user is presented with a progress bar while the code executes. The length of time to complete the calculation depend on the amount of data available for a reach and can range from less than a minute to several minutes for very complex reaches. After the code completes execution, the user can click the "Download Results" button to get all the results in a single compressed file (zip format) containing each of the results files for the target reach as images or tab-delimited text files.
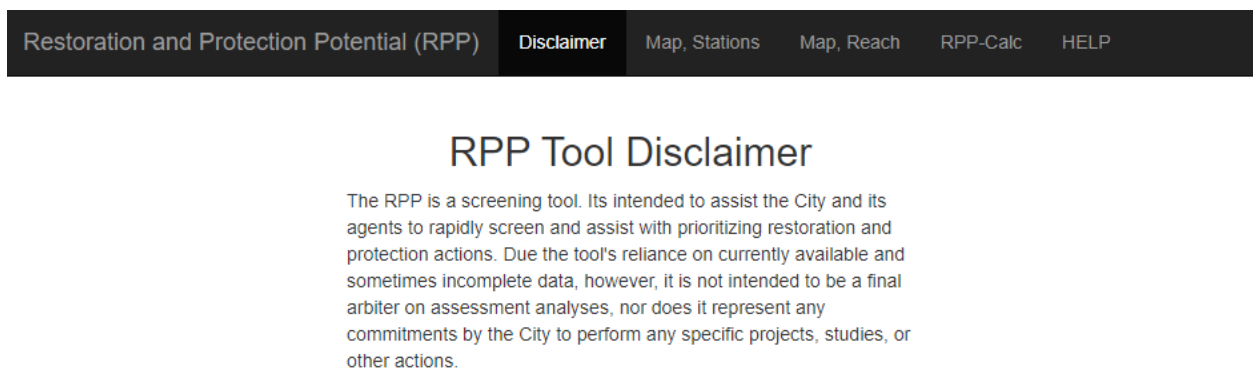
Figure 5-2. Shiny app start screen with disclaimer
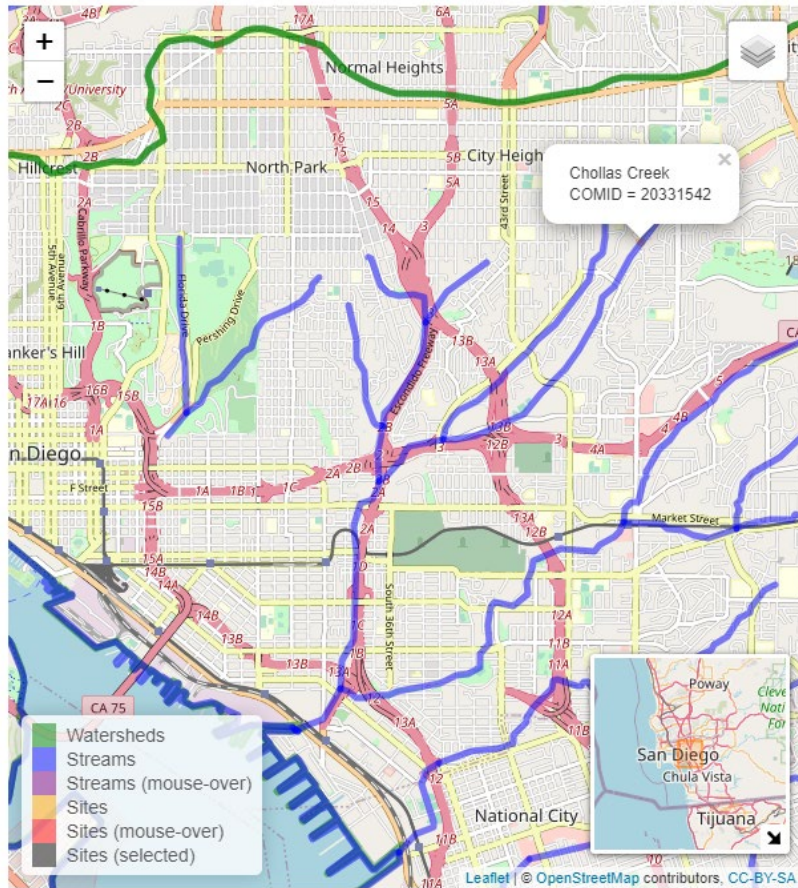
Figure 5-3. Shiny app map for locating sites.
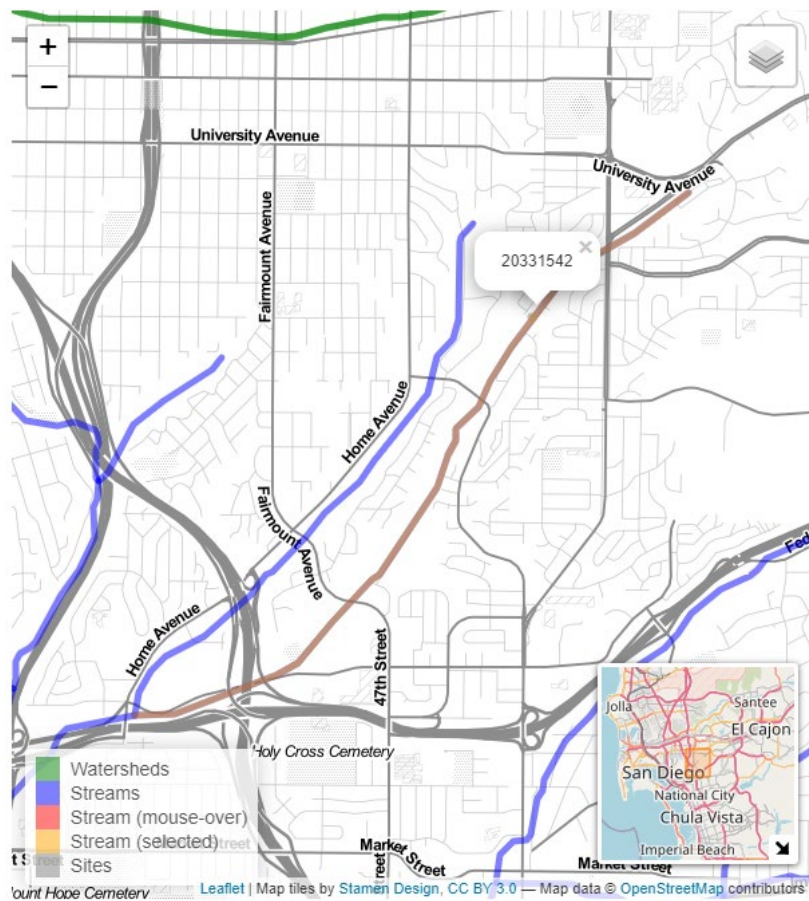
Figure 5-4. Shiny app map for locating stream reaches with alternate base map.

Figure 5-5. Shiny app running tool.

6.0 References

Beck, M. W., Mazor, R. D., Johnson, S., Wisenbaker, K., Westfall, J., Ode, P. R., . . . Stein, E. D. (2019). Prioritizing management goals for stream biological integrity within the developed landscape context. *Freshwater Science, 38*(4), 883-898. doi:10.1086/705996.

CalFire. (2007). *Fire Hazard Severity Zones Maps*. Sacramento, CA: State of California, Office of the State Marshal. Retrieved from https://osfm.fire.ca.gov/divisions/wildfire-planning-engineering/wildland-hazards-building-codes/fire-hazard-severity-zones-maps/. Accessed 17 March 2020.

California Department of Fish and Wildlife. (2020). *NCCP Plan Summary - San Diego Multiple Species Conservation Program (MSCP)*. Sacramento, CA: California Department of Fish and Wildlife, Habitat Conservation Planning Branch. Retrieved from https://wildlife.ca.gov/Conservation/Planning/NCCP/Plans/San-Diego-MSCP.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2019). Shiny: Web Application Framework for R.R. Package Version 1.3.2. Retrieved from https://CRAN.R-project.org/package=shiny.

City of San Diego, Planning and Development Services. (1997). Multiple Species Conservation Program (MSCP). Retrieved from https://www.sandiegocounty.gov/pds/mscp/. Accessed 29 March 2020.

Gillett, D. (2018, 12 January). [SCCWRP Regional Boundary].

Mazor, R. D., Rehn, A. C., Ode, P. R., Engeln, M., Schiff, K. C., Stein, E. D., . . . Hawkins, C. P. (2016). Bioassessment in complex environments: designing an index for consistent meaning in different settings. *Freshwater Science, 35*(1), 249-271. doi:10.1086/684130.

Ode, P. R., Rehn, A. C., Mazor, R. D., Schiff, K. C., Stein, E. D., May, J. T., . . . Hawkins, C. P. (2016). Evaluating the adequacy of a reference-site pool for ecological assessments in environmentally complex regions. *Freshwater Science, 35*(1), 237-248. doi:10.1086/684003.

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

SANDAG/SanGIS. (2015). *SanGIS/SANDAG GIS Data Warehouse*. San Diego Association of Governments (SANDAG) and San Diego Geographic Information Source (SanGIS),. Retrieved from https://rdw.sandag.org

SANDAG/SanGIS. (2018a). SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal - Existing Landuse. Retrieved from https://sdgis-sandag.opendata.arcgis.com/datasets/existing-landuse-1. Accessed 21 March 2020.

SANDAG/SanGIS. (2018b). SANDAG/SanGIS Regional GIS Data Warehouse Open Data Portal - Planned Landuse. Retrieved from https://sdgis-sandag.opendata.arcgis.com/datasets/planned-landuse-1. Accessed 21 March 2020.

SCCWRP Staff (2019). [SMC Portal: Biological community sample data from 1997-2017].

Tetra Tech, Inc. (2019). *Causal Assessment Screening Tool for the Supplemental Environmental Project (SEP) Support Restoring and Protecting Our Waterways: Bio-assessment Tools and Priorities for Action*. San Diego, CA: City of San Diego.

The City of San Diego. (1998). *Multiple Species Conservation Program (MSCP Plan)*. San Diego, CA: The City of San Diego. Retrieved from https://www.sandiegocounty.gov/content/dam/sdc/pds/mscp/docs/SCMSCP/FinalMSCPProgramPlan.pdf.

USEPA. (2016). *Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems* (EPA/822F-16/002). Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://www.epa.gov/wqc/practitioners-guide-biological-condition-gradient-framework-describe-incremental-change-aquatic.

USEPA. (2017). *Causal Analysis/Diagnosis Decision Information System (CADDIS)*. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development. Retrieved from https://www.epa.gov/caddis.

USEPA. (2018). *EnviroAtlas - Native Freshwater Aquatic Biodiversity by 12-digit HUC for the Conterminous United States*. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory. Retrieved from https://www.epa.gov/enviroatlas/forms/enviroatlas-data-download. Accessed 16 April 2020.

USEPA. (2019). *EnviroAtlas: Native Aquatic Species Vulnerability Index (NASVI)*. Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://enviroatlas.epa.gov/enviroatlas/DataFactSheets/pdf/ESN/Nativeaquaticspeciesvulnerabilityindex.pdf.

USGS and USEPA. (2012). National Hydrography Dataset Plus (NHDPlus Version 2. Medium Resolution for California). Retrieved from https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus. Accessed 11 November 2012.

Appendix C. RPPTool Prioritized Reaches

Pilot Watersheds Combined

This appendix presents the top-ranked, potential restoration and protection reaches within the pilot watersheds that were used to help guide development of the tools (Chollas, Los Peñasquitos, San Diego River). The results are preliminary and represent the default setup to help support further analysis and application of these tools to inform future management actions. Note there were no top-ranked reaches within the Chollas watershed.
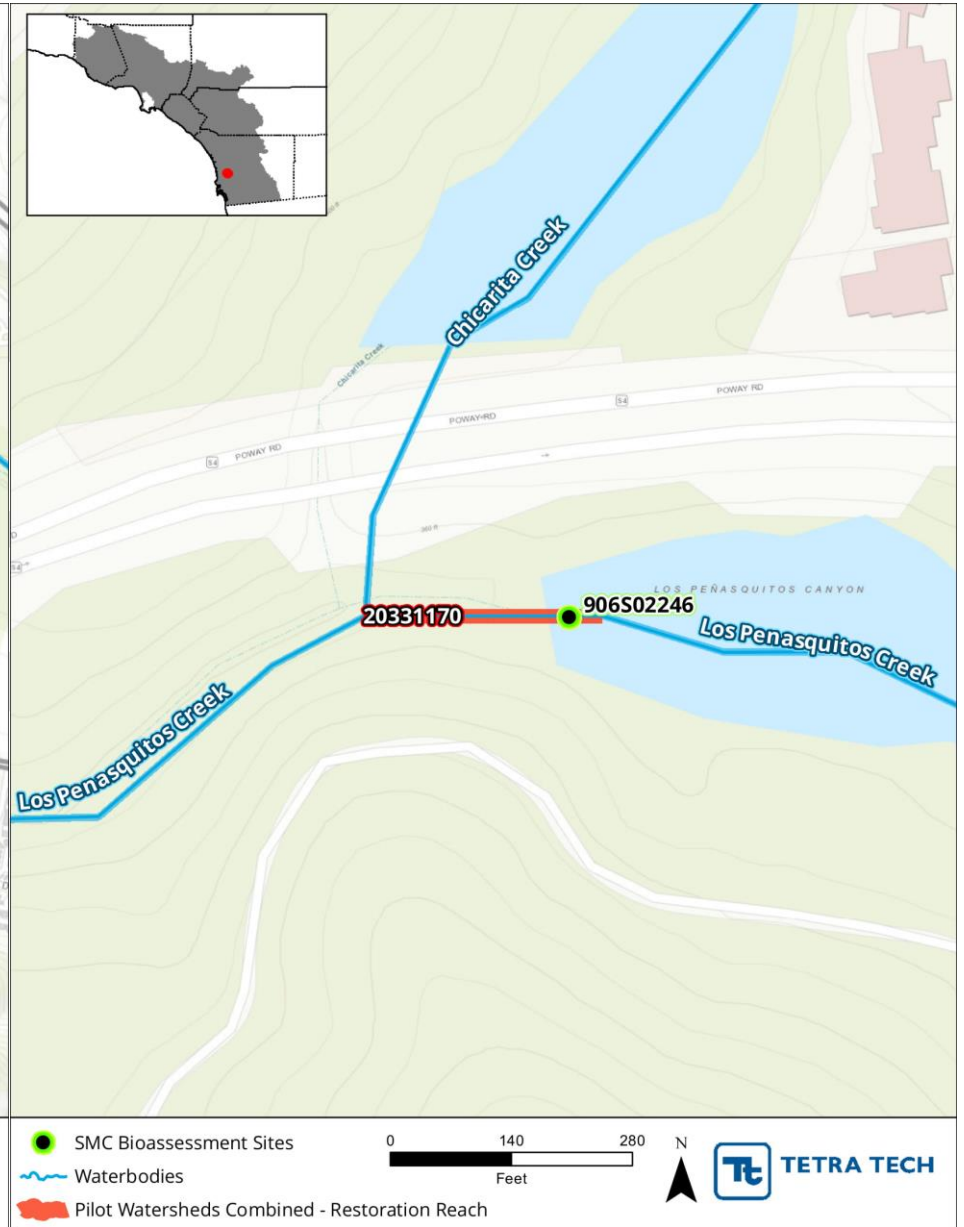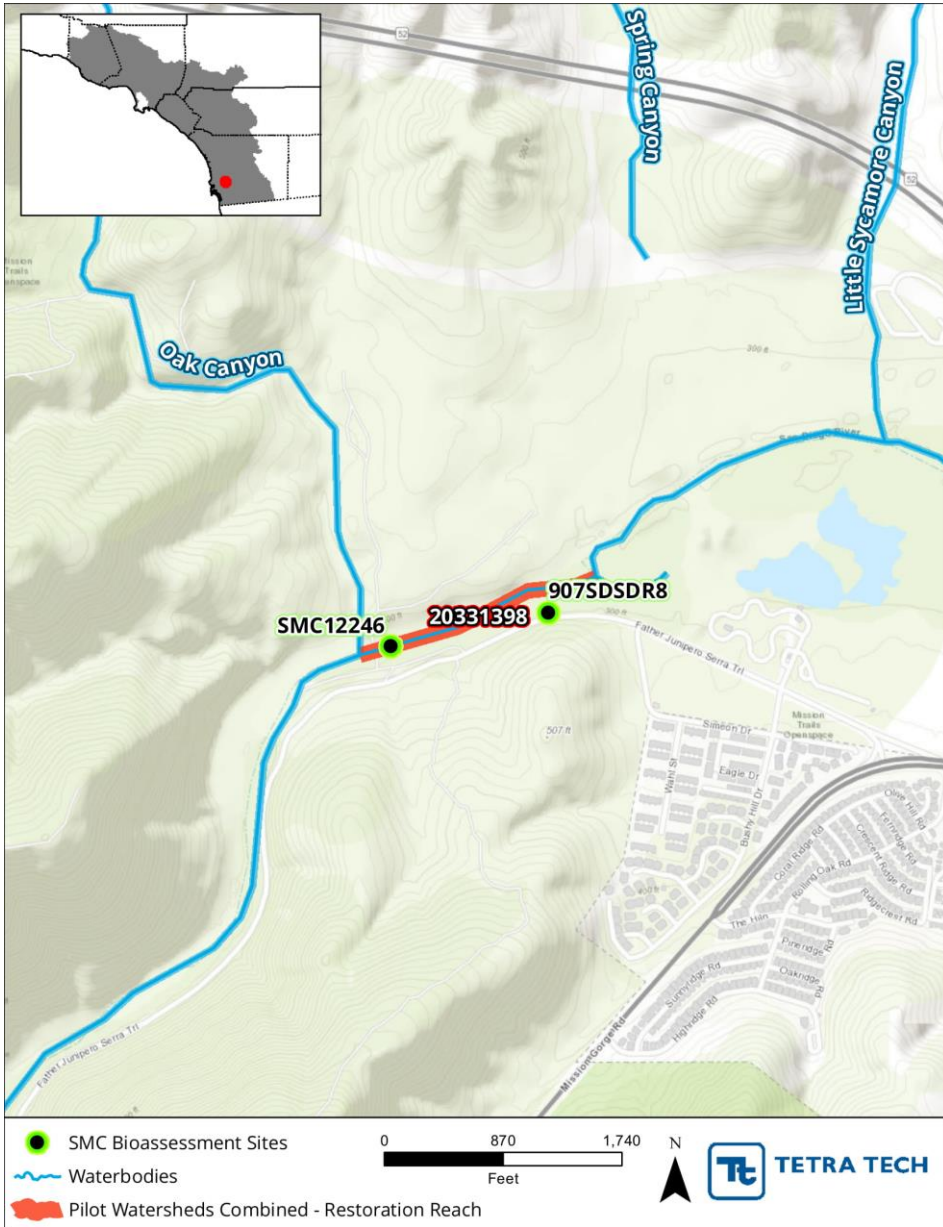
The RPPTool options for user-defined co-benefits and weighting provide opportunities to consider other factors such as socio-economic incentives and synergies with other resource management opportunities. Future iterations of restoration and protection ranking can incorporate user-defined criteria in scoring and supplement existing processes and plans such as the City of San Diego's Watershed Asset Management Plan (WAMP), Watershed Master Plans (WMPs), and Water Quality Improvement Plans (WQIPs) to evaluate locations and approaches for restoration and protection prioritization.
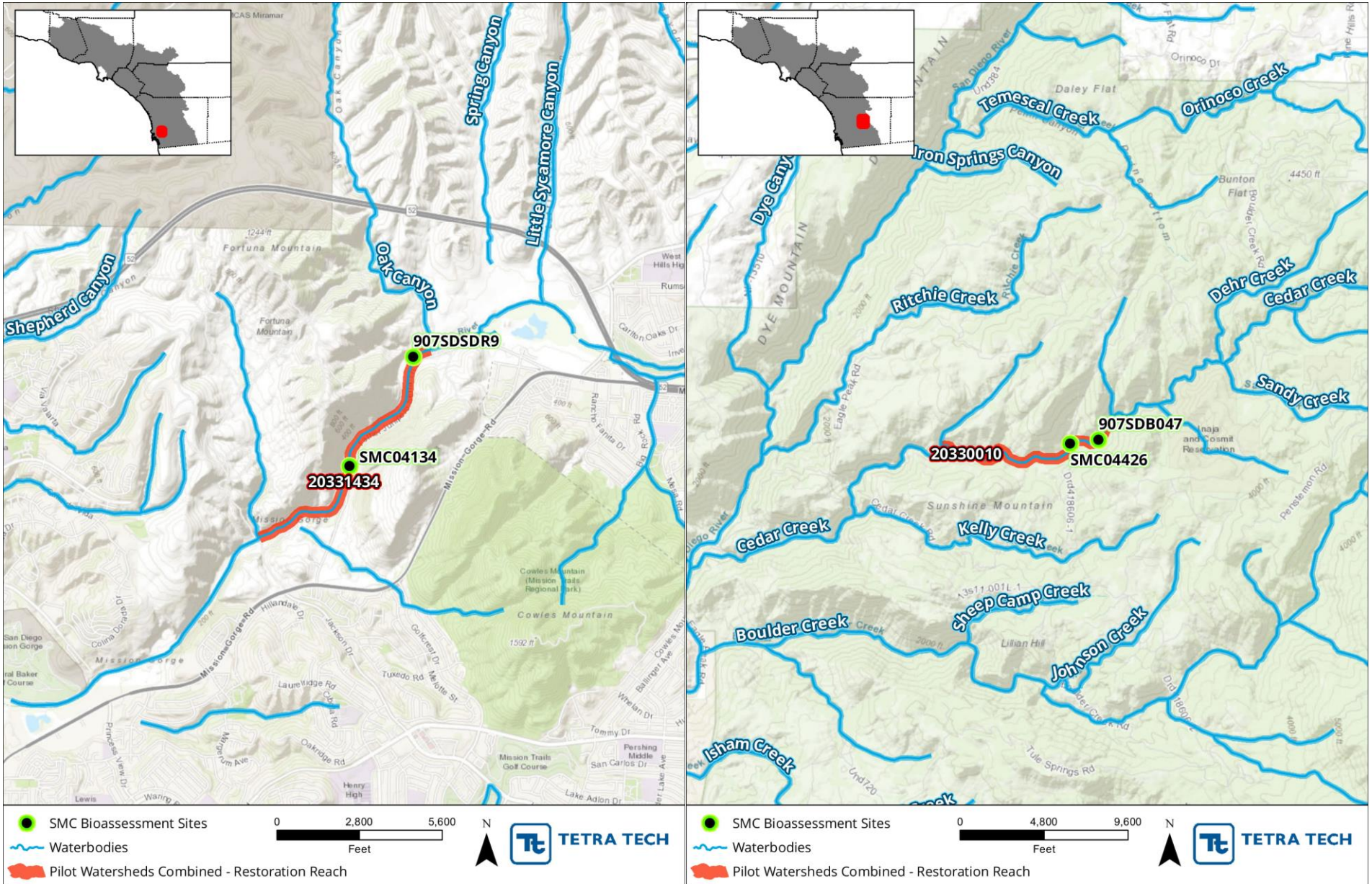
As a pilot project intended to inform the development of the tools, the sites listed in the tables below are not slated for restoration or protection based on this effort. The RPPTool index ranking represents one type of assessment the City may use in its planning process. The City uses multiple resources and plans to prioritize sites for restoration and protection and the RPPTool offers additional information that can be considered in future decision making.
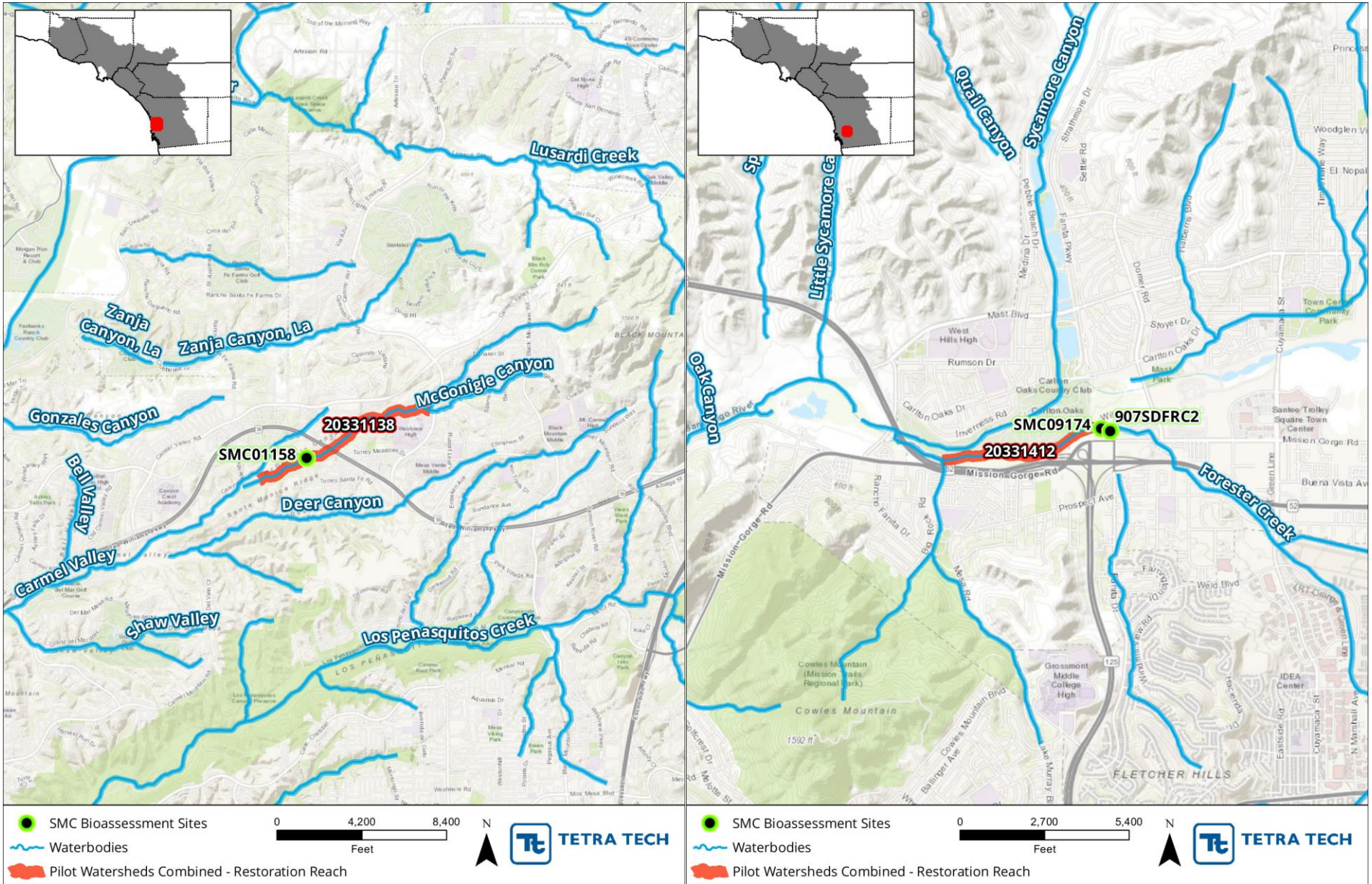
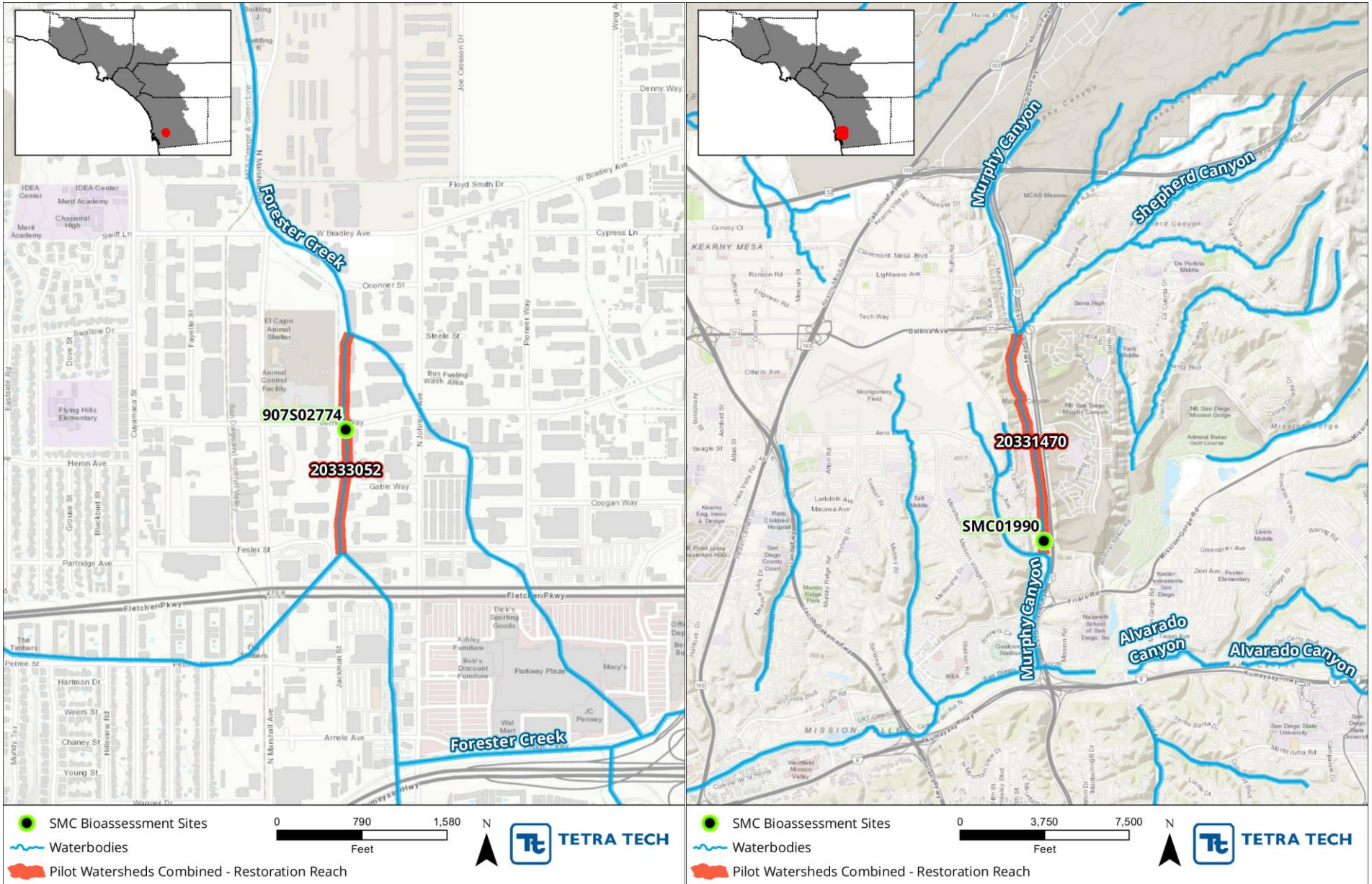**Table 1. Top 10 Restoration Reaches for Pilot Watersheds**

Rank	COMID	StationID	Watershed	CSCI	BCGTier	RPPIndex	Notes
1	20331398	SMC12246	San Diego River	0.41	5	0.476	See page C-2
2	20331170	906S02246	Los Peñasquitos	0.77	4	0.45	See page C-2
3	20331434	SMC04134	San Diego River	0.78	3	0.45	See page C-3
4	20330010	SMC04426	San Diego River	0.7	4	0.447	See page C-3
5	20331138	SMC01158	Los Peñasquitos	0.5	5	0.447	See page C-4
6	20331412	SMC09174	San Diego River	0.54	5	0.441	See page C-4
7	20333052	907S02774	San Diego River	0.69	4	0.433	See page C-5
8	20331470	SMC01990	San Diego River	0.66	4	0.427	See page C-5
9	20331318	SMC04806	Los Peñasquitos	0.78	3	0.426	See page C-6
10	20331934	SMC03110	San Diego River	0.7	4	0.423	See page C-6

The City of SAN DIEGO

Left map:
Spring Canyon
Little Sycamore Canyon
Oak Canyon
907SDSDR8
SMC12246 20331398
Father Junipero Serra Trl

SMC Bioassessment Sites
Waterbodies
Pilot Watersheds Combined - Restoration Reach

0   870   1,740
Feet
N
TETRA TECH

Right map:
Chicarita Creek
Poway Rd
LOS PEÑASQUITOS CANYON
20331170   906S02246
Los Penasquitos Creek
Los Penasquitos Creek

SMC Bioassessment Sites
Waterbodies
Pilot Watersheds Combined - Restoration Reach

0   140   280
Feet
N
TETRA TECH

The City of
SAN DIEGO

Legend (left map):
- ● SMC Bioassessment Sites
- ～ Waterbodies
- ▬ Pilot Watersheds Combined - Restoration Reach

0   790   1,580
Feet

N

Tt TETRA TECH

Site labels (left map): 907S02774, 20333052, Forester Creek

Legend (right map):
- ● SMC Bioassessment Sites
- ～ Waterbodies
- ▬ Pilot Watersheds Combined - Restoration Reach

0   3,750   7,500
Feet

N

Tt TETRA TECH

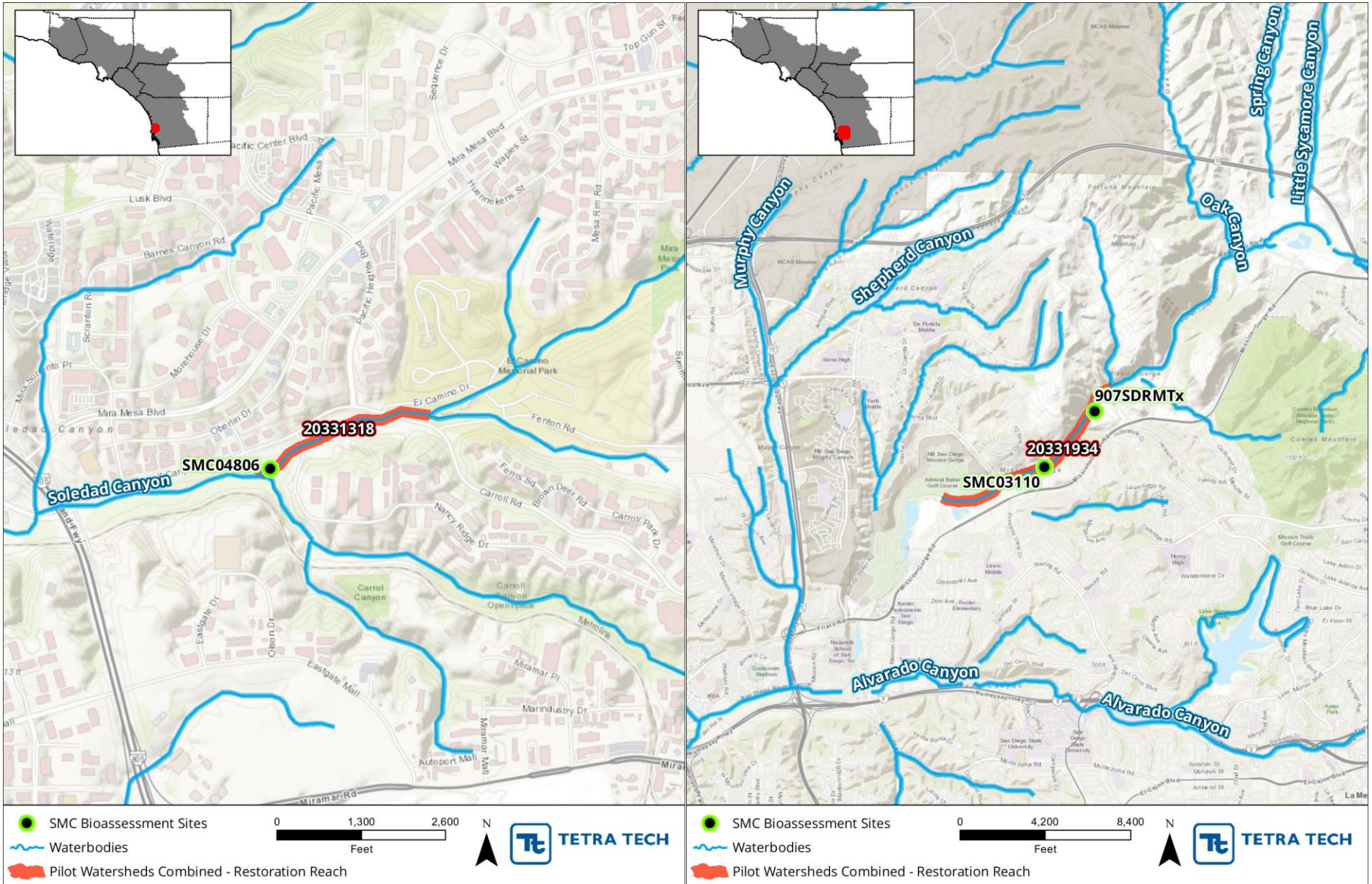Site labels (right map): Murphy Canyon, Shepherd Canyon, 20331470, SMC01990, Alvarado Canyon

Table 2. Top 10 Protection Reaches for Pilot Watersheds

| Rank	COMID	StationID	Watershed	CSCI	BCGTier	RPPIndex	Notes
1	20330038	907S03210	San Diego River	1	2	0.461	See page C-8
2	20332438	907S01418	San Diego River	1.22	2	0.448	See page C-8
3	20331208	SMC04294	Los Peñasquitos	0.87	3	0.446	See page C-9
4	20332528	SMC04682	San Diego River	1.19	2	0.445	See page C-9
5	20329834	907S00577	San Diego River	1.06	2	0.438	See page C-10
6	20332496	907S05514	San Diego River	1.12	2	0.432	See page C-10
7	20332830	907S03786	San Diego River	1.22	2	0.419	See page C-11
8	20332828	SMC32718	San Diego River	0.82	3	0.417	See page C-11
9	20334398	907S46499	San Diego River	1.07	2	0.414	See page C-12
10	20332754	907S01434	San Diego River	0.98	3	0.412	See page C-12
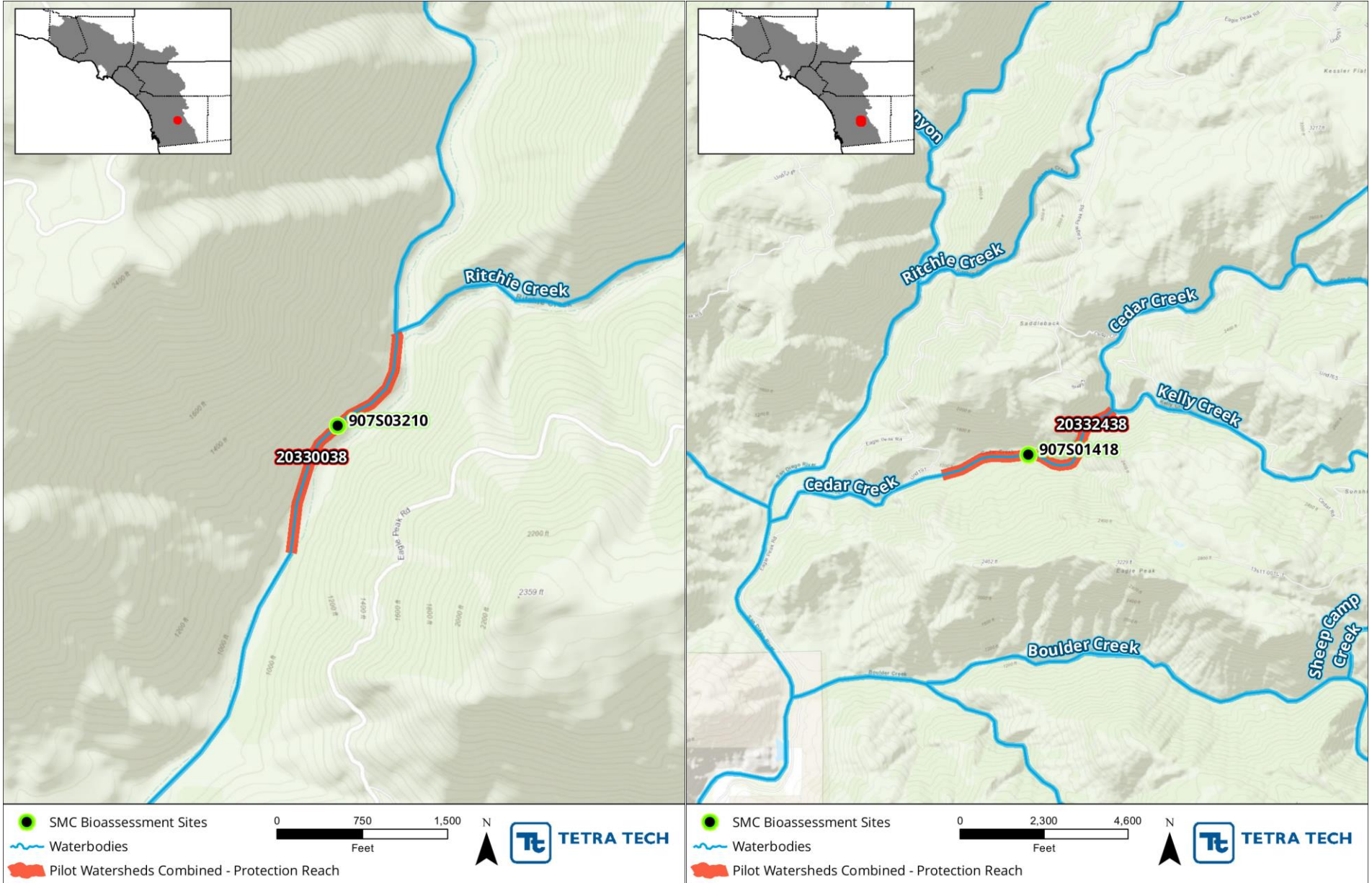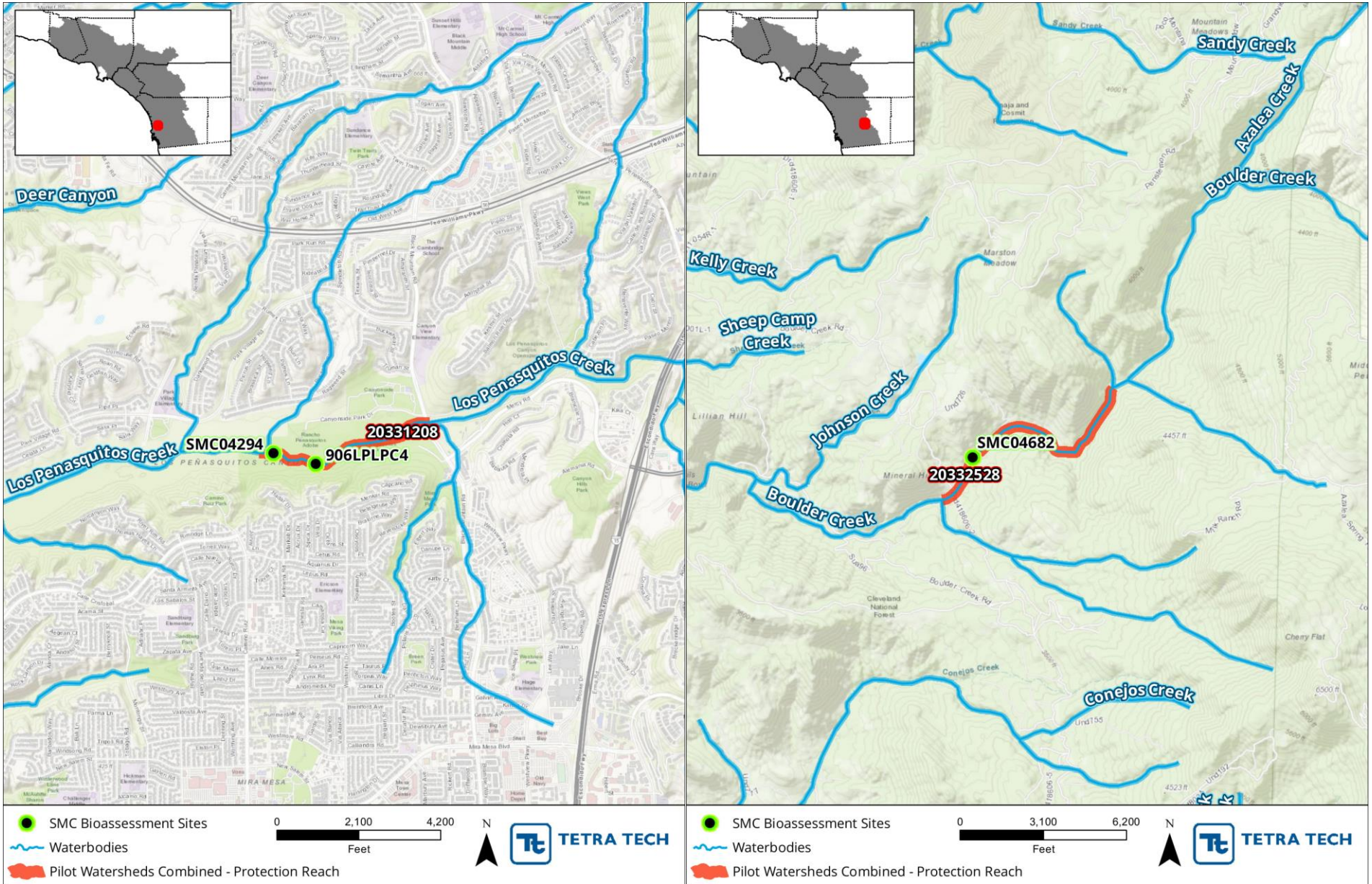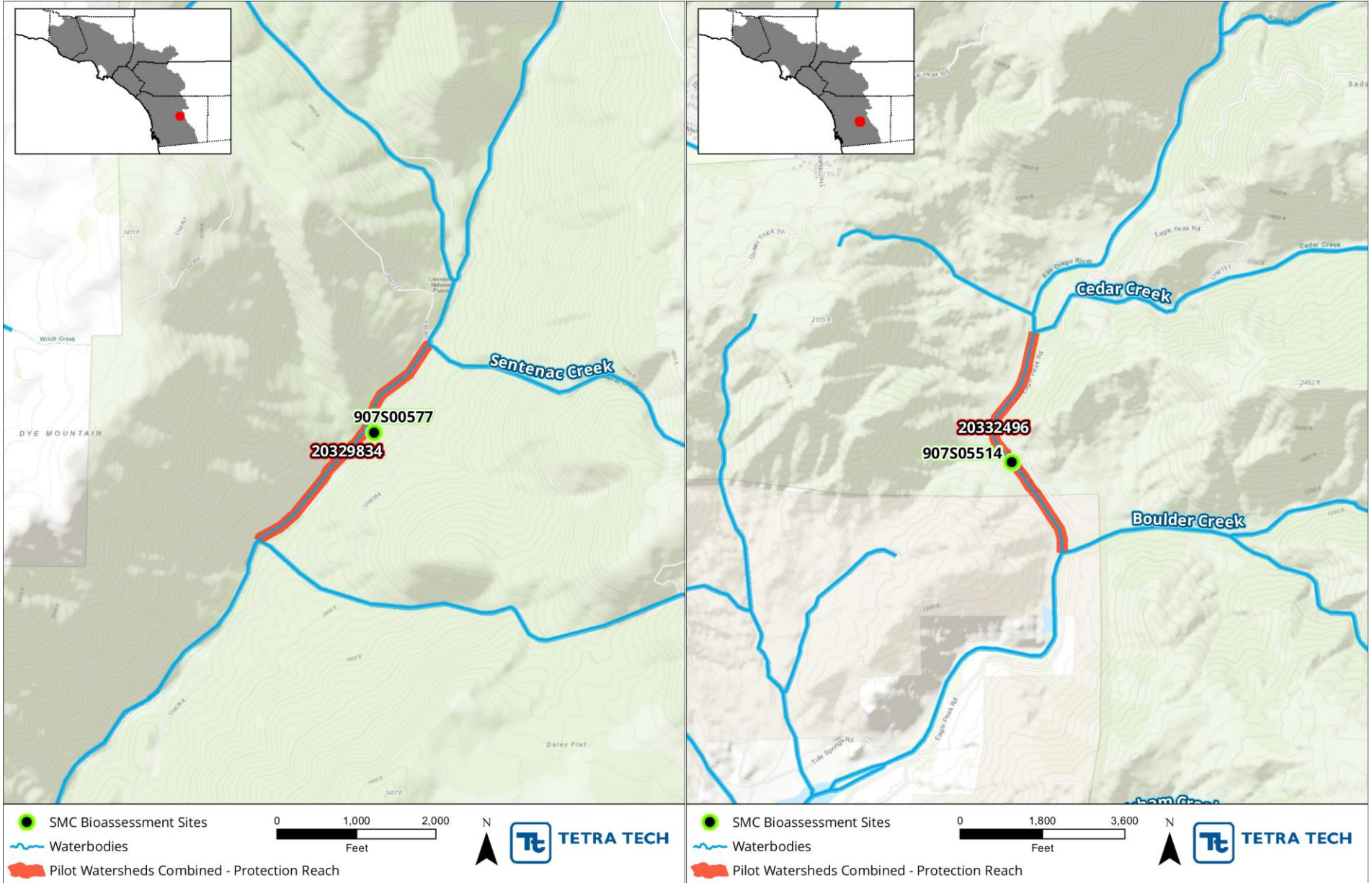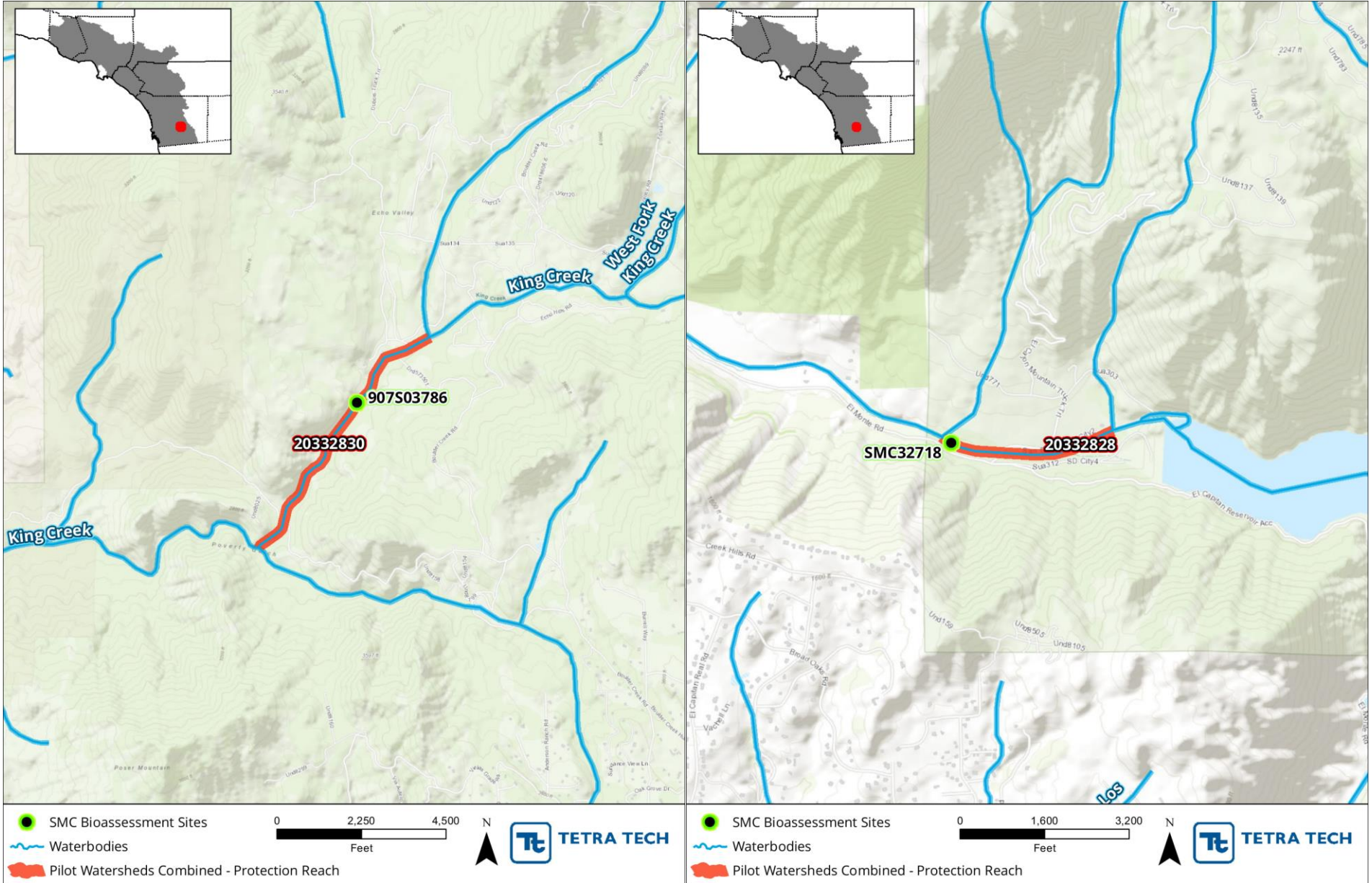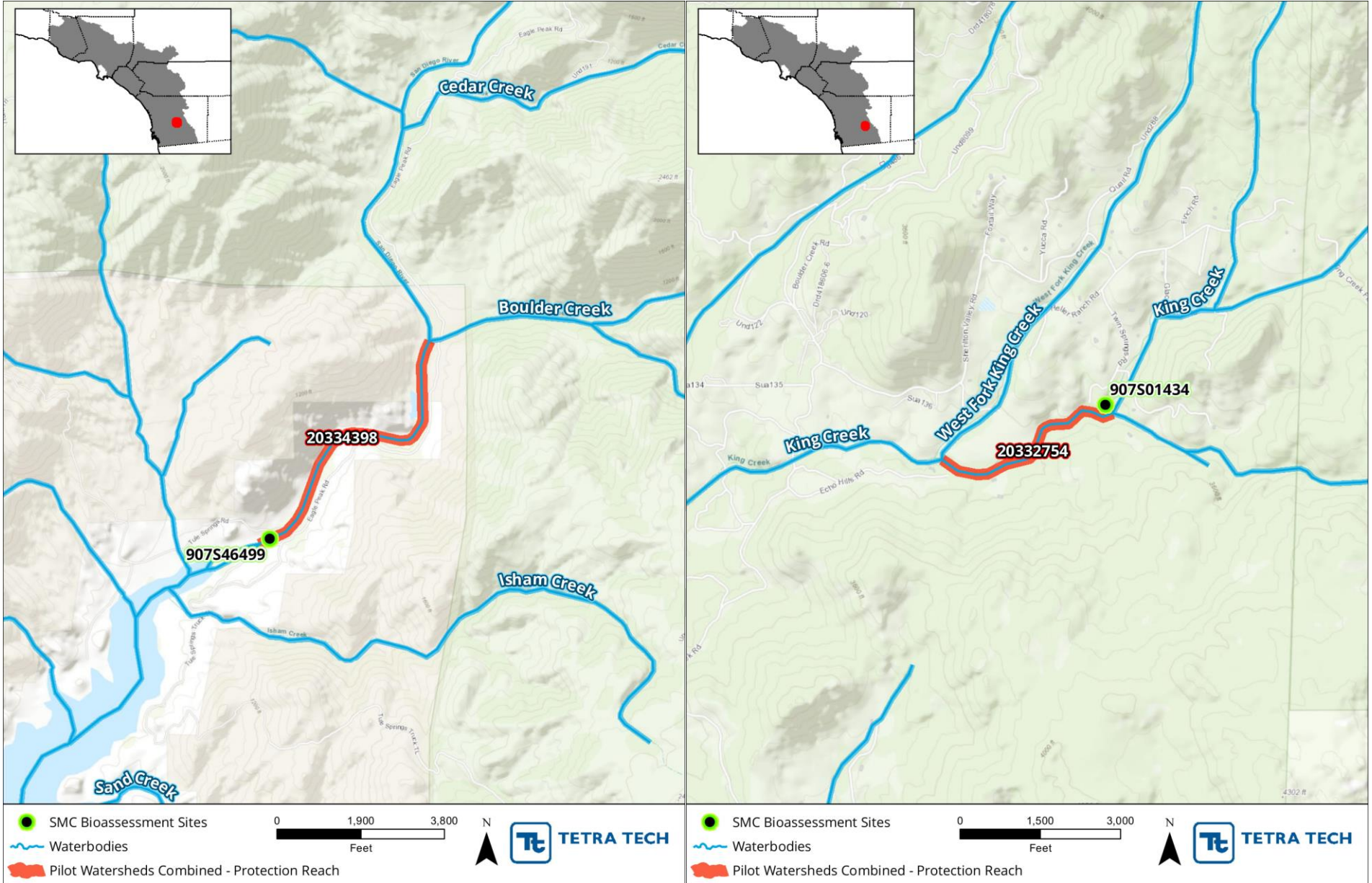
Appendix D. Outreach

Outreach Overview

Outreach activities conducted as part of the Supplemental Environmental Project (SEP) included several ongoing and completed strategies to reach target audiences, describe project activities, and provide access to the RPPTool and CASTools. In accordance with the SEP provisions, the outreach materials describe the project as being undertaken as part of a settlement of San Diego Water Board enforcement action. A wide range of audiences have been engaged through previously conducted and ongoing outreach activities. The outreach activities include:

- Eight presentations to stakeholder groups, scientists, regulators, and water quality managers on regional, state, national, and international levels;
- Two project fact sheets developed to illustrate the CASTool and RPPTool functionalities and potential uses;
- Emails to stakeholders in the three priority watersheds: Los Peñasquitos, Chollas (San Diego Bay), and San Diego River;
- City of San Diego website highlight; and
- Social media postings

Summary of Outreach Activities

Presentations

The presentations detailed in Table D-1 were given during the course of the SEP to a broad range of audiences from regional stakeholders to international experts.

Table D-1. Summary of Presentations Conducted During the Project

Date	Entity	Title
May 2017	Society of Environmental Toxicology and Chemistry (SETAC) Europe	A Causal Assessment Approach that Rapidly Evaluates Multiple Stressors in Stream Systems
Sept. 2017	California Stormwater Quality Association (CASQA)	Screening Causal Assessment to Support Development of Biological Objectives and Prioritizing Implementation Actions
Oct. 2018	California Biologists Workshop	A Causal Assessment Approach that Rapidly Evaluates Multiple Stressors in Stream Systems
Nov. 2018	SETAC North America Conference	A Methodology to Predict Toxicity Potential of an Urban Stormwater Discharge
		Comparator Site Selection to Inform Screening-Level Causal Assessments

Date	Entity	Title
Feb. 2019	California State Water Resources Control Board (SWRCB)	SEP Support Restoring and Protecting Our Waterways: Bioassessment Tools and Priorities for Action
Sept. 2020	CASQA	Development of a Stream Restoration and Protection Prioritization Tool
Mar. 2021	Southern California Stormwater Monitoring Coalition	CASTool and RPPTool Overview

Project Fact Sheets

Fact sheets were developed for the RPPTool and CASTool. The fact sheets provide project details in an easily distributable format with content accessible to multiple audiences. The sheets are included as Attachment D-1 and links to the fact sheets are available on the City's website.
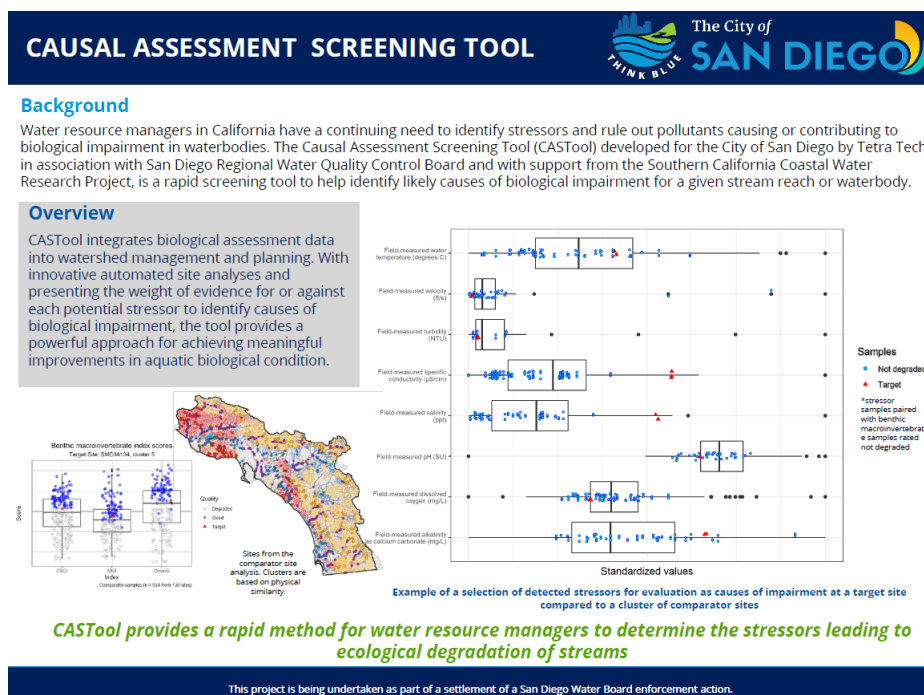

Figure D-1. Screenshot of the CASTool Fact Sheet.

Website

The City of San Diego's Pilot Project website has been updated to include project information. The website will be used in social media posts and emails to stakeholders to direct them to more information. The website currently includes a short description about the project, links to the Fact Sheets, and the 2020

CASQA presentation. In the future, links to the Shiny app versions of the tools and the final report will be added to the page. The page is available through the following link: https://www.sandiego.gov/stormwater/pilot-projects and a screenshot of the current page is included below.
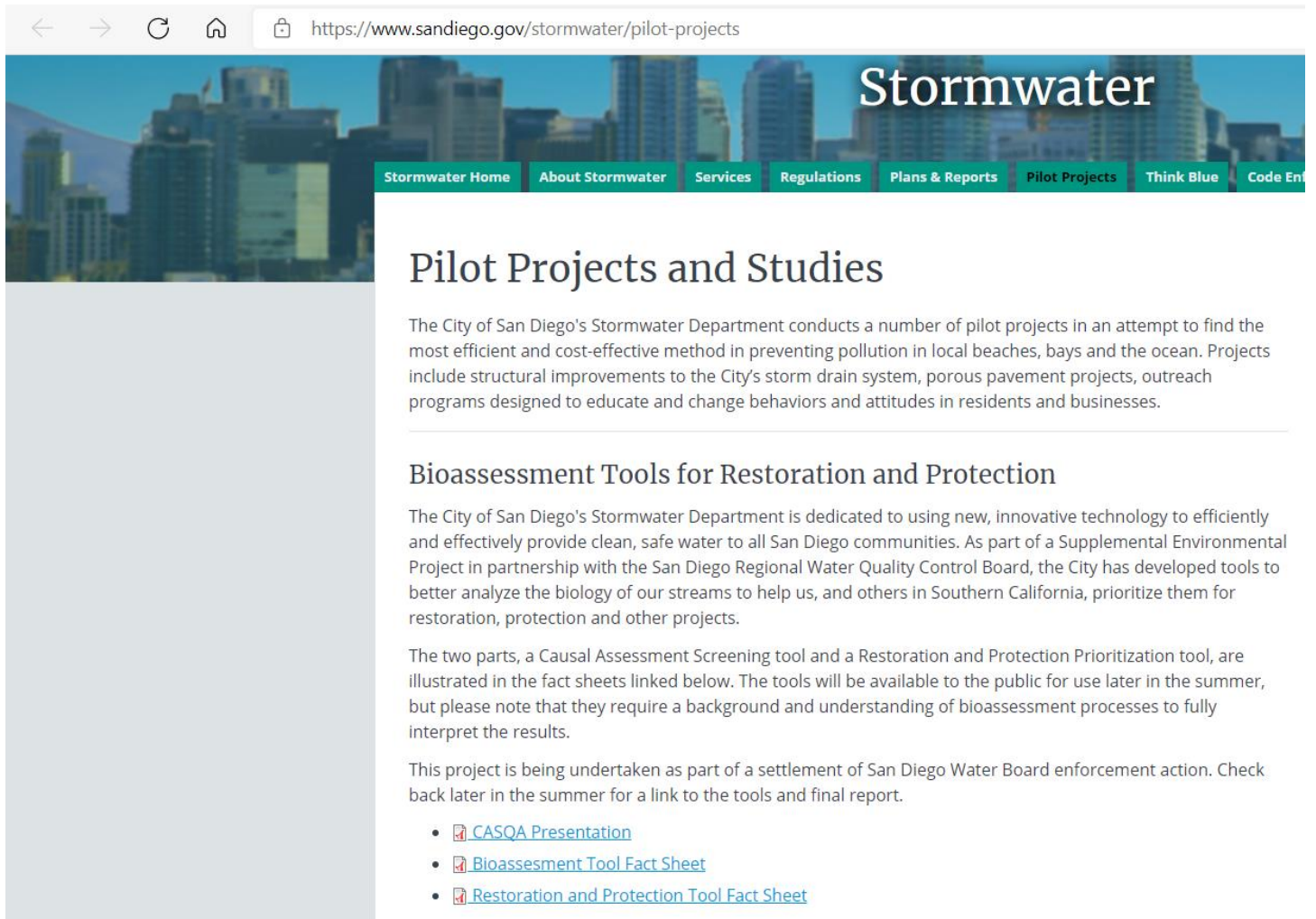
Figure D-2. Screenshot of SEP Description and Project Links on the City's Website.

Stakeholder Emails

To reach the stakeholders of the priority watersheds Los Peñasquitos, Chollas (San Diego Bay), and San Diego River, the following text will be distributed to Consultation Committee members via email.

Hello Consultation Committee Members,

The City of San Diego's Stormwater Department is dedicated to using new, innovative technology to efficiently and effectively provide clean, safe water to all San Diego communities. The Department has been working on some new tools to better analyze the biology of our streams to help us prioritize them for restoration, protection and other projects.

We have recently developed a Bioassessment Tool (Tool) as part of a Supplemental Environmental Project with the San Diego Regional Water Quality Control Board. This tool features two parts, which are outlined in the attached fact sheets. These will be available to the public for use later in the summer, but please note that they require a background and understanding of bioassessment processes to fully interpret the results. This project is being undertaken as part of a settlement of a San Diego Water Board enforcement action. You can find more information about this tool and how it can be used on our website https://www.sandiego.gov/stormwater/pilot-projects. Check back later this summer for links to the tools and our final reports detailing how the tool was developed.

Thank you and please reach out to Vicki Kalkirtz at vkalkirtz@sandiego.gov with any questions.

Figure D-3. Consultation Committee Email Text

Social Media

Social media postings are planned to provide direct outreach to the public. Postings will include brief content and the website link above for users to access more information. The City's Think Blue Facebook page is a planned forum for posting.
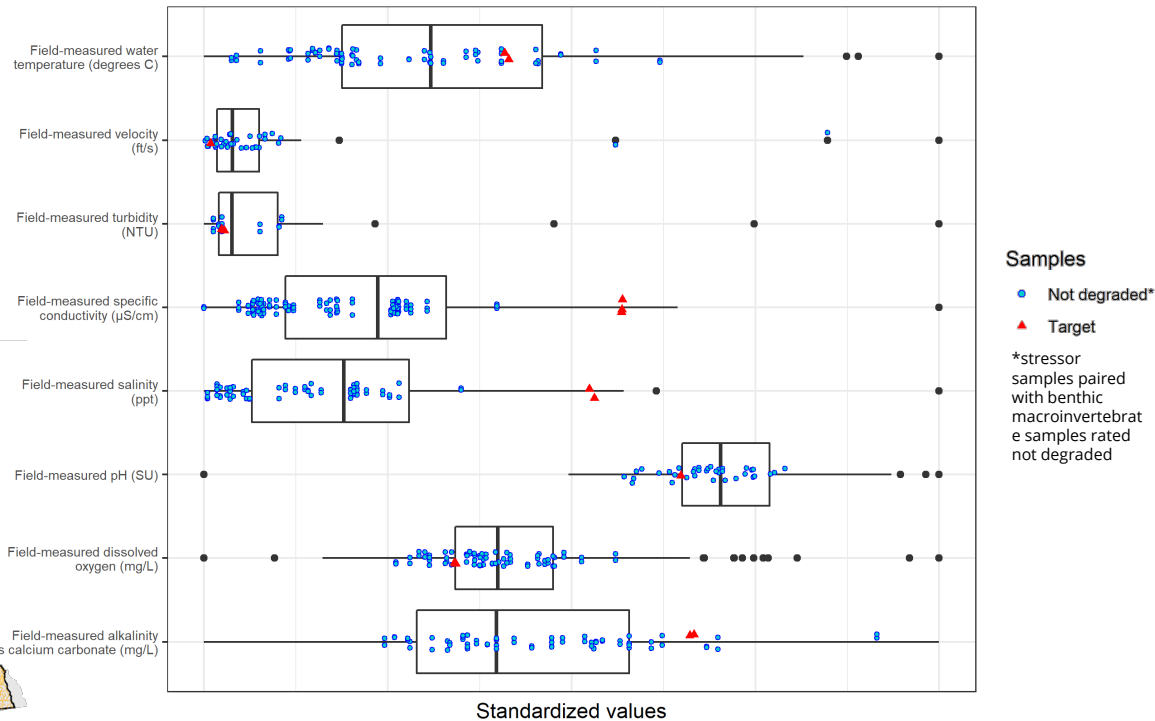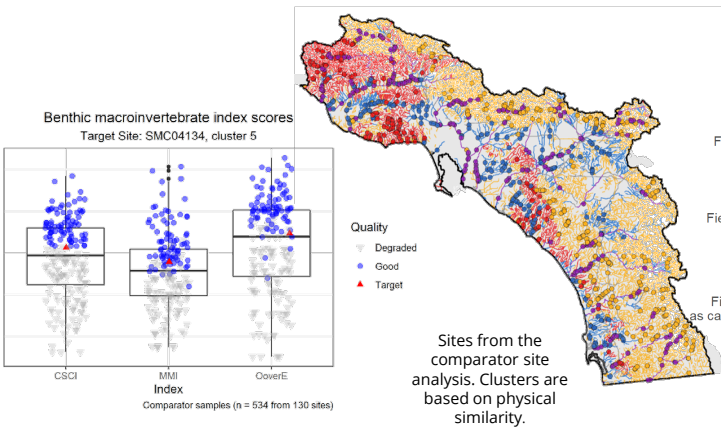
Attachment D-1. CASTool and RPPTool Fact Sheets

Background

Water resource managers in California have a continuing need to identify stressors and rule out pollutants causing or contributing to biological impairment in waterbodies. The Causal Assessment Screening Tool (CASTool) developed for the City of San Diego by Tetra Tech in association with San Diego Regional Water Quality Control Board and with support from the Southern California Coastal Water Research Project, is a rapid screening tool to help identify likely causes of biological impairment for a given stream reach or waterbody.
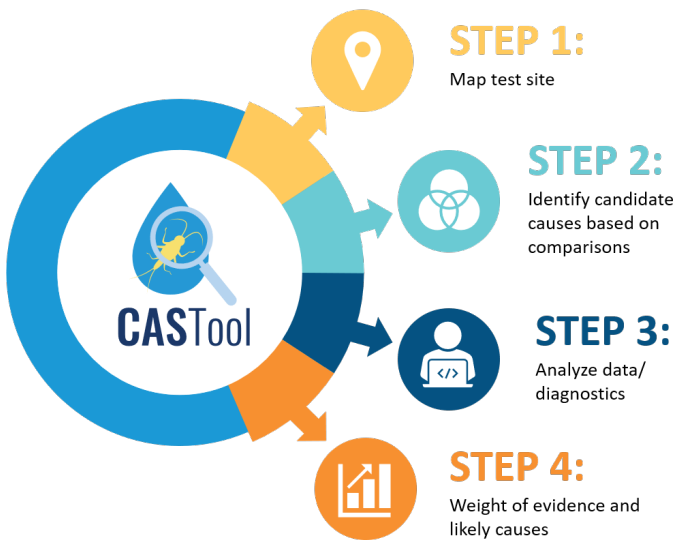
Overview

CASTool integrates biological assessment data into watershed management and planning. With innovative automated site analyses and presenting the weight of evidence for or against each potential stressor to identify causes of biological impairment, the tool provides a powerful approach for achieving meaningful improvements in aquatic biological condition.

Benthic macroinvertebrate index scores
Target Site: SMC04134, cluster 5

Quality
▽ Degraded
● Good
▲ Target

Sites from the comparator site analysis. Clusters are based on physical similarity.

Samples
● Not degraded*
▲ Target

*stressor samples paired with benthic macroinvertebrate samples rated not degraded

Standardized values

Example of a selection of detected stressors for evaluation as causes of impairment at a target site compared to a cluster of comparator sites

CASTool provides a rapid method for water resource managers to determine the stressors leading to ecological degradation of streams

This project is being undertaken as part of a settlement of a San Diego Water Board enforcement action.

RAPID IDENTIFICATION OF THE LIKELY CAUSES OF BIOLOGICAL IMPAIRMENT

The City of SAN DIEGO
THINK BLUE

STEP 1: Map test site

STEP 2: Identify candidate causes based on comparisons

STEP 3: Analyze data/ diagnostics

STEP 4: Weight of evidence and likely causes

CASTool

How it Works

Data

CASTool uses stressor data such as water quality, physical habitat metrics/indices, and flow metrics. Biological response data include the California Stream Condition Index (CSCI) for benthic macroinvertebrates, the Algal Stream Condition Index (ASCI) and their respective subindices and metrics.

Candidate Causes (Stressors)

The likelihood of potential stressors as causes of observed biological impairment at a site are ranked. This allows for determination of the relative importance of each candidate cause using a weight of evidence approach.

User Interface

CASTool is available via a user-friendly web-application framework built using Shiny. A more advanced application and enhancements to allow for additional user flexibility may be pursued in the future.

## *CASTool incorporates both algal and macroinvertebrate data and identifies likely stressors*

Lines and Weight of Evidence

CASTool evaluates specific lines of evidence that are frequently used in traditional causal assessments (e.g., EPA's CADDIS) and provide the strongest evidence as to whether a stressor is likely to be a cause of concern at a site. Graphical results along with displays of the input data are produced.

Scores from each line of evidence are incorporated in a Weight of Evidence analysis. Transparent criteria are used within the tool to determine whether results support or refute a stressor as a likely cause of the observed biological condition.

Example of a Stressor-Response line of evidence
Is there evidence of a biological gradient from inside or outside the case?
Linear regression with 75th percentile prediction interval

Samples
- all
- not degraded*
- comparator
- comparator not degraded*
- target
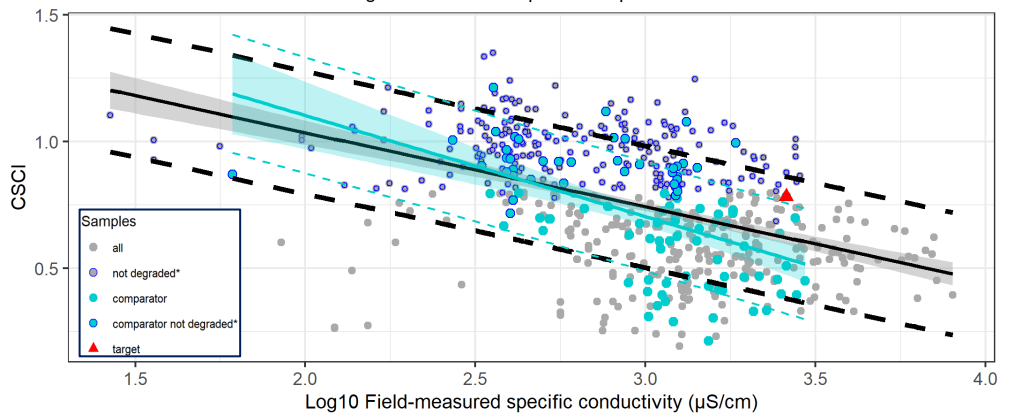
Regression (all, outside the case): y = -0.293 x + 1.62 ~ r2 = 0.220 ~ p-value = 2.6e-32 ~ n = 566 ~ score = 1
Regression (comparators, inside the case): y = -0.399 x + 1.9 ~ r2 = 0.274 ~ p-value = 1.4e-08 ~ n = 103 ~ score = 1

Project Contacts

Vicki Kalkirtz
City of San Diego
vkalkirtz@sandiego.gov

Clint Boschen
Tetra Tech
clint.boschen@tetratech.com