Exploratory categorization of watersheds for potential stormwater monitoring in San Francisco Bay

Ben Greenfield, Marcus Klatt, and Lester McKee, SFEI[a]

Summary

This document presents technical information regarding watershed classification schemes and potential sampling priorities. Though this information is needed to develop appropriate study designs, the optimal study design to address the MRP questions must be based on a combination of this technical information, monitoring and management considerations, and stakeholder priorities. Thus, the final decision regarding appropriate study design is pending stakeholder input.

The purpose of this document is to summarize efforts to categorize watersheds based on available information. The 185 watersheds in the study area (Figure 1) contain a variety of attributes available in GIS layers. Using this information, cluster analysis is performed to categorize the watersheds into a manageable number of categories (8 categories is selected). Two alternative categorization schemes are also presented, based on combinations of simple attributes that exhibited strong gradients in the data set. The attributes used were percent imperviousness in combination with either PG&E facilities or historic railroads.

Introduction

The San Francisco Bay Regional Water Quality Control Board (Water Board) has developed Total Maximum Daily Load reports (TMDLs) for Hg and PCBs (SFBRWQCB 2006, 2008). These TMDLs summarize available knowledge, provide linkages between waste loads and beneficial uses, and prescribe mass load reductions aimed at bringing San Francisco Bay into compliance with water quality objectives or other applicable standards (SFBRWQCB 2006, 2008). Both TMDLs call for increased effort by stormwater agencies to manage and reduce loads over a 20 year period (2028 for Hg and 2030 for PCBs). The TMDLs allow for wasteload allocations of 82 kg of Hg and 2 kg of PCBs in urban stormwater. These represent estimated reductions of 50% and 90% over the present load estimates of 160 kg of Hg and 20 kg of PCBs. However, the current loads estimates are highly uncertain. In addition, since one method of demonstrating compliance is to determine trends in loads (either mass or particle concentrations), there is a need for increased effort to measure loads.

This need is reflected in the recently adopted municipal regional stormwater NPDES permit (MRP) (SFBRWQCB 2009) that covers the co-permittees of the cities of Vallejo,

[a] We thank Michelle Lent, Aroon Melwani, Arleen Feng and Chris Sommers for project guidance and direction. Jamie Kass provided GIS technical assistance. Kat Ridolfi, Alicia Gilbreath, and Sarah Pearce assisted with data compilation. Michelle Lent and Arleen Feng reviewed drafts of this document. This project was funded by the Regional Monitoring Program as part of the Sources Pathways and Loadings Work Group.

Fairfield, and Suisun, and the counties of Contra Costa, Alameda, Santa Clara, and San Mateo. Provision C.8.e of the MRP calls for pollutants of concern (POCs) monitoring that is intended to assess inputs of POCs to the Bay from local tributaries and urban runoff. This monitoring should provide a basis to assess progress toward achieving wasteload allocations (WLAs) for TMDLs. It is also intended to help resolve uncertainties associated with loading estimates for these pollutants. Consistent with this permit requirement, the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP) has developed a Small Tributaries Loadings Strategy (STLS). The STLS is intended to help RMP and Bay Area Stormwater Management Agencies Association (BASMAA) efforts synergistically achieve common objectives laid out by the permit. The STLS and the permit provision C.8.e. were developed in parallel and contain the same basic management questions:

1) Identify which Bay tributaries (including stormwater conveyances) contribute most to Bay impairment from pollutants of concern
2) Quantify annual loads or concentrations of pollutants of concern from tributaries to the Bay;
3) Quantify the decadal-scale loading or concentration trends of pollutants of concern from small tributaries to the Bay; and
4) Quantify the projected impacts of management actions (including control measures) on tributaries and identifying where these management actions should be implemented to have the greatest beneficial impact.

A long-standing recommendation of the Sources Pathways and Loadings Workgroup (SPLWG), a subgroup of the RMP, is to stratify watersheds into general categories and then to sample a subset of watersheds in selected categories. Two key questions in relation to the STLS and the MRP, are: 1. how many types of watersheds occur in the permitted region and, 2. how many watersheds should be studied to answer key management questions? In response to needs of the MRP as well as the RMP (described below), this study has two main objectives:

1) To develop and document a rationale for initially classifying Bay Area small tributary watersheds into a small number (<10) of classes, relevant for loads monitoring and Bay margin impacts
2) To provide STLS stakeholders with a tool to help develop a list of representative watersheds in each class, and rank them for focused follow-up evaluation.

Categorizing watersheds and determining which watersheds to study will provide a basis for improving the cost effectiveness of developing loads information. It will also provide support for other strategies and initiatives. For example, the RMP *Modeling Strategy* is developing a Bay margins conceptual model that will benefit from establishing priority watersheds, and from compiling watershed contaminant and process data. In addition, the RMP Small Fish Study (Greenfield and Jahn 2010), originally conceived in the *Mercury Strategy,* is proposed to be expanded to include PCBs in 2010. Information developed for classifying watersheds could help to identify candidate sites for small fish sampling. As the RMP continues to support process studies for contaminant uptake into the food web, this watershed classification study will help identify "high leverage" areas on the Bay margin likely to have relatively large food web impacts.

Ordination and cluster analysis are exploratory techniques designed to visualize patterns on complex multivariate data sets. Cluster analysis is particularly aimed at identifying unique groupings (i.e., Clusters) within the data set, based on the combined differences of multiple attributes. In ecology, cluster analysis is frequently performed to generate categories of habitats, or other groupings of sampling events, based on overall patterns of species abundance and distribution. Cluster analysis has also been used to categorize and characterize water bodies and watersheds, based on general land cover attributes available through GIS (Eilers et al. 1983, Young and Stoddard 1996, Bulley et al. 2007). In the present exercise, we use this technique to categorize watersheds in sections of San Francisco Bay, based on available land use, land cover, and other environmental data. The intent is to form a basis for developing a sampling scheme to evaluate contaminant and suspended sediment loading in relation to the TMDLs and MRP permit provision C.8.

Methods

Watershed boundary delineation

In order to generate statistical characteristics of Bay Area watersheds as basic input data for a cluster analysis, a series of spatial data layers were retrieved from local, state and federal agencies. Central to this analysis was a watershed boundary layer, the modern form of which has been in development for several decades. The challenge with urban watershed boundaries is that much of the drainage system is underground and only loosely follows topographic landscape features. Since the early 1990s the RMP, BASMAA, and Bay Area flood control agencies have been involved in an effort to collate a geographic information system (GIS) map of the urban drainage infrastructure for the Bay Area at a regional scale. At present, watershed boundaries have been generated for most watersheds in western Contra Costa County, Alameda County, Santa Clara County, and San Mateo County and are downloadable from a range of locations on the internet including SFEI (http://www.sfei.org/projects/3051). Additional watershed boundary information was provided from the Contra Costa County Watershed Atlas for eastern Contra Costa County. The area of the present study includes watersheds in these counties, for which high quality watershed boundary information is readily available (Figure 1).

Bayland portions of the watersheds with tidal influence were removed from this analysis using the EcoAtlas Modern Bayland Habitat layer. Some of the 185 watersheds have available subwatershed data, but were aggregated for this exploratory analysis. However, areas upstream of major dams were treated as separate "subwatersheds" in this analysis, as these were assumed to trap all sediments and associated contaminants and therefore not be valid as future study locations for assessment of contaminant loads impacting San Francisco Bay. Therefore, these areas will not be selected for sampling, regardless of the statistical analysis output. A point file from the National Inventory of Dams was sorted by the size of the drainage into the reservoir. All points with a drainage area over 20 square miles (approximately 50 square kilometers) were considered consistent with the previous work of Davis et al. (2000), resulting in the removal of the areas above 10 major dams in the study area. The upstream portions of the watersheds were digitized using the 10 meter Digital Elevation model (DEM), 10 meter DEM hillshade, USGS Topoquads, and the National Hydrologic Dataset (NHD) flow lines including the South Bay storm drains. Area was then calculated for each watershed polygon. The total area delineated upstream of these points is 1,597 km^2.

The resulting GIS boundary layer shape file includes 185 watersheds ranging in size from 0.023 to 962 km^2 and covering a total area of the total 5,630.5 km^2 in the counties of Contra Costa, Alameda, Santa Clara, and San Mateo. Watersheds in the jurisdictions of Fairfield and Suisun were not included because high quality data have not yet been assembled on watershed boundaries and other attributes.

GIS watershed attributes

Statistics for each watershed were generated about population, various land use types including modern and historical industrial area and rail transport lines, areas of greater likelihood of PCBs and Hg contamination (for example auto-wrecking yards), areas of greater PCB use (e.g. PG&E facilities), and other relevant layers thought to be useful for classifying watersheds in relation to PCB and Hg loading studies (McKee et al. 2006). Each of these layers, and their basis for inclusion, are described in more detail below.

Population was calculated from the 2000 Census Block Groups shapefile. The Block Group polygons were split by the watersheds. The resulting split population polygon populations were recalculated by multiplying the area of the split polygon by the population per unit area. These populations were then summed within each watershed.

Land use was calculated from the ABAG 1995 Regional Existing Land Use dataset (ABAG 1997). The land use polygons were split by watershed. The land use categories were generalized into industrial, residential, commercial, open space and agriculture. “Null” land use values were excluded from analysis. These null values were generally Bayland features, upland reservoirs, and portions of the watershed outside the boundaries of the land use dataset. The land use data were then dissolved by watershed and type. The area per watershed of each land use type was attributed to each watershed.

PCBs and Hg are classified as legacy contaminants. Although small amounts are still in use today, the peak use of both substances occurred more than three decades ago. Both substances were used in industrial applications; consequently, soils and sediments in historic industrial areas are often contaminated (van Geen and Luoma 1999, Kuzyk et al. 2005b, SFBRWQCB 2006, 2008). A historic industrial land use dataset is based on land use that is classified "urban" in 1954 (USGS) reference maps and "industrial" in current reference maps (ABAG 2000). The 1954 reference maps do not distinguish industrial land uses from other urban uses (e.g. residential, commercial). In order to estimate historic industrial land use, we made the assumption that any areas that are currently industrial land use that intersect with the historic urban land use layer were historically industrial. This was assumed because it is unlikely that residential, commercial, agricultural or open space land use would be converted to industrial land use given trends in the Bay Area are dominantly towards urban residential and commercial land uses. This data layer was intersected by watershed and summed by area.

The volume of runoff that occurs in urban areas is influenced by the area of impervious surfaces. Since PCBs and Hg are predominantly transported into San Francsico Bay during rain storms and stormwater runoff (McKee et al. 2005), permeability may be a good indicator of PCB and Hg loads. Permeability was calculated by converting the NLCD 2001 Impervious layer (NLCD 2001) to polygons and dissolving by percent permeable surfaces. This polygon layer was then intersected with the watershed layer. This polygons area was calculated for each permeability value polygon and this area was multiplied by the percent permeability to create the amount of impervious surface. The amount of impervious surface was then summed by watershed.

Soils in areas around railroads have been identified as having greater concentrations of PCBs and Hg (McKee et al. 2005). This is probably due to a variety of reasons including incidental spillage during loading and transport, the use of both PCBs and Hg in electrical applications such as switching and motive power, and the use of used industrial oils for dust suppression (McKee et al. 2006). Railroad data layers were compiled separately for current vs. historic railroads again focusing on the period of greater PCBs and Hg use (1950-1990). The current railroad layer was created by the USGS as part of the digital line graphic database. The historic railroads layer was created using rectified 1951 to 1961 USGS topographic quads. The rail lines that were not included in the current-day rail lines were digitized using the heads-up methodology (i.e., directly traced on the computer screen using scanned raster images as a backdrop). Each railroad data layer was intersected by watershed and then the length was summed per watershed. A total rail length per watershed was created by summing these fields.

Both PCBs and Hg were used heavily in the auto industry. For example, PCBs were used in electrical starters, capacitors, and as flame retardants in upholstery. Hg was used in electrical components including switches, thermostats, and halogen lights. Area in the urban landscape where vehicles are recycled, refurbished or disposed of are likely to be subjected to contamination. An auto dismantlers data layer was created representing the active auto and truck dismantling facilities (i.e., auto wreckers and junk yards) listed in the San Francisco Regional Water Quality Control Board's records in October 2002. These facilities can be a source of ground and surface water contamination and are thus closely monitored by the Regional Board. The facility locations have been determined by address-match geocoding supplemented by hand-plotting using aerial photographs and maps, but the data has not been error-corrected. This data layer was intersected by watershed and then the count was attributed to each watershed.

The largest use of PCBs (60% in the US) was in the power generation and transmission industry. USEPA PCB self-reporting data base (http://www.epa.gov/epawaste/hazard/tsd/pcbs/pubs/data.htm) about 260,000 kg

(580,000 lb) of PCBs are currently being cycled out of use each area in California. While PG&E is cleaning up their facilities in compliance with all laws and regulations, legacy contamination in soils is still present on PG&E properties at concentrations that while legal, may be slowly being dispersed off site by wind, water, wheel- and foot-tracking and entering the local stormwater conveyance (SFEI, 2010). These areas are known to have concentrations greater than TMDL targets (McKee et al. 2006). PG&E Facilities data were compiled from PG&E database for Bay Area obtained by the San Francisco Regional Water Quality Control Board (RWQCB) by Fred Hetzel in 2002. These data were intersected by watershed and then the count was attributed to each watershed.

Both PCBs and Hg are transported into the Bay predominantly associated with sediment particles. As a result, the RMP make considerable annual effort to measure or improve estimates of suspended sediment loads entering the Bay (McKee et al. 2005, McKee et al. 2006, David et al. 2009, Lewicki and McKee 2009). Recently, suspended sediment load estimates were generated for small tributaries in Bay Area (Lewicki and McKee 2009). These were obtained and used here.

Many urban watersheds in the Bay region drain areas on the Bay margin that are near or even below sea level or where stormwater, on its way to the Bay, must pass below infrastructure such as freeways and railways. These physio- and socio-graphic features in the urban landscape have necessitated the use of pumps to lift stormwater out to the Bay. These areas are often current or historical industrial areas and are often also in close proximity to wastewater treatment facilities. Provisions C.11.f. and C.12.f. of the MRP call for permittees to implement five pilot projects to divert dry weather and first flush flows to wastewater treatment plants to address these flows as a source of PCBs and Hg to receiving waters. While there are other reasons, it is primarily the proximal relationship between industrial land uses, pump stations, and wastewater treatment facilities that make this option seem attractive. For these reasons, we chose to include information on pump stations as a factor of influence for future watershed loads monitoring. The Pump Stations data layer was developed in a collaborative effort of SFEI, BASMAA, and the Water Board. The Water Board requested information from Phase 1 permittees during the fall of 2007 and Caltrans in 2009. SFEI organized the information into a database and GIS shapefile in March of 2009. These data were intersected by watershed and then the count was attributed to each watershed.

Precipitation is a major driver in the transport of POCs to San Francisco Bay. For example it have been estimated that >99.5% of Hg loads entering the Bay from the Guadalupe River watershed during an average year do so during the wet season (McKee et al. 2005). A precipitation layer was obtained from http://frap.cdf.ca.gov/ . This layer represents lines of equal rainfall (isohyets) based on long-term mean annual precipitation data. These data were collected over a sixty year period (1900-1960), over a minimum mapping unit of 1000+ acres. Data were compiled from USGS, California Department of Water Resources, and California Division of Mines map and information sources. Source maps are based primarily on U.S. Weather Service data for approximately 800 precipitation stations, and supplemented by county and local agency precipitation data. These data were intersected by watershed, and then the average rainfall was calculated by summing the area of each rainfall value polygon multiplied by the area and dividing it by the total area.

The history of urbanization and industrial land use in the Bay Area has lead to residues of PCBs and Hg in urban soils and in the sediments of the stormwater conveyance system. Knowledge about the distribution of soil and sediment contamination may provide a basis for estimating which watersheds may have greater loads. Consequently, over the past 10 years, BASMAA and SFEI have been gradually increasing information on soil and sediment contaminant concentrations. Presently, Hg and/or PCB concentrations are reported for over 700 data points disbursed throughout the counties of San Francisco, San Mateo, Santa Clara, Alameda, Contra Costa and Solano (no data points are located in the counties of Marin, Napa, or Sonoma). This watershed PCB and Hg in sediment point data layer includes data from multiple studies regarding the concentrations of pollutants in street and storm drain sediments around the San Francisco Bay Area (City of San Jose and EOA 2003; EOA, Inc. 2004; EOA Inc. 2007; EOA 2002, Gunther et al. 2001; KLI and EOA 2002; Kleinfelder 2005; Kleinfelder 2006; Solop, Abu-saba et al 2002; Solop, Harding et al 2002). The most recent version of this dataset was compiled as part of the San Francisco Estuary Institute's "Regional Stormwater Monitoring and Urban BMP Evaluation"http://sfei.org/stormwaterbmps/gis_data/Soil_dust_drop_inlet_sediment_Hg_PCB_concentrations.zip. Each point was intersected and averaged by watershed.

A premise of the STLS is the ability to identify tributaries having controllable sources that exert a disproportionately large influence on loads and impacts on the Bay margin. As the RMP continues to support process studies for contaminant uptake into the food web, this watershed classification study will help identify “high leverage” areas on the Bay margin likely to have relatively large food web impacts. Given these broader objectives and synergies between RMP strategies, we included knowledge of Bay margin characteristics in the present data compilation. A within-Bay sediment Hg data layer includes Hg data from multiple studies regarding the concentration of hg in San Francisco Bay sediments. This dataset (N = 699 points total) was compiled in 2009 (Greenfield et al. 2009) with additional data added in February of 2010. Data sources include studies by the RMP, CALFED, USGS, the South Bay Mercury Project, the PRISM Program, and other unpublished data sets. Each point was spatially joined with the nearest individual watershed within a 500 meter search radius limit. A total of 98 of the 700 points were attributed to specific watersheds.

Data processing

All parameters were scaled to watershed area except UTM northing and easting spatial coordinates and precipitation. Thus, for example, land cover parameters were converted into proportion of total area. The % water land cover was removed from the analysis, as 182 of 185 results were zeros, and the remaining 3 values were less than 0.1% of total land cover. The two historical railroad results were summed into a single historic railroad category. Also removed from the statistical analysis were parameters that were not available for all watersheds: subtidal sediment Hg, watershed soil Hg, and model predicted watershed suspended sediment loading.

To improve results of distance calculation measurements, multivariate normality and linearity, and consistent scales among the parameters is desirable. To achieve this, parameters were transformed; proportions were arcsin(sqrt) transformed[b], population was sqrt transformed, and other parameters were log transformed. Precipitation and geospatial coordinates (UTM northing and easting) were approximately normally distributed, and therefore not transformed. To maintain the presence of zero values in the data set, log transformations were adjusted following the procedure of zero conversion described in Chapter 9 of McCune and Grace (2002). All results were then scaled from 0 to 1. Examination of bivariate scatter plots and histograms confirmed approximate multivariate normality and linearity.

Statistical analysis

Two dissimilarity measures were applied to the converted data set: relative Euclidian dissimilarity (i.e., standardized to the sum of squares to achieve maximum values of sqrt(2)) and Bray-Curtis dissimilarity (i.e., Sørensen index). Hierarchical clustering was performed on each dissimilarity matrix using Ward's minimum variance method as the linkage method (McCune and Grace 2002). Other methods were attempted (average linkage, single linkage, and complete linkage) but dendograms indicated poorly defined clusters with substantial chaining (addition of single items to existing groups). Clustering was also performed using the "clues" algorithm, which uses an automated combination of partitioning and shrinking to optimize cluster number (Chang et al. 2010). The clues method was implemented with a Euclidian distance dissimilarity measurement, and the Silhouette index to optimize cluster number. Silhouette and CH indices were evaluated to compare the clustering methods. Silhouette index compares the dissimilarity of each point within clusters to all of the points in the nearest neighboring cluster. CH index is the ratio of variation within clusters vs. variation between clusters (Chang et al. 2010).

The data set was also evaluated using Non-Metric Multidimensional Scaling, performed with the metaMDS algorithm (Minchin 1987, Cox and Cox 1994, 2001), and based on Bray-Curtis dissimilarity. The NMDS was performed assuming two axes. MetaMDS uses random starting configurations to avoid local optima and find the global best solution. A convergent solution was not found after 20 iterations, when comparing solutions using Procrustes rotation, suggesting that a stable solution may not exist, and warranting caution in interpretation. However, all solutions resulted in highly similar stress. Statistical analyses were performed in R (version 2.10.1), using the vegan, clues, and MASS packages.

[b] Arcsin(sqrt) transformations are commonly performed to improve normality of proportion or percentage data.

Results and Discussion

Cluster analysis results

Selection of number of clusters is user dependent, and based on the intent of the analysis. Cluster analysis performed with the clues algorithm indicated that seven clusters was the optimal number. For this analysis, we are trying to differentiate the 185 watersheds into a tractable number of categories. Given this objectives of the study, the results of the clues algorithm, results of dendograms using the different clustering methods (e.g., Figures 2 and 3), and the attributes of resulting clusters, eight categories was selected as the appropriate number for interpretation.

Silhouette and CH indices were similar among the three clustering methods, suggesting that they exhibited a similar ability to partition the data set into unique, compact, and dissimilar clusters. Measures of similarity among cluster results only indicated moderate correspondence among the methods. This was corroborated by graphical analysis of NMDS results (presented later), and suggests that the successful partitioning of this data set is not robust to different clustering methods.

Graphical analyses of box and whiskers plots were performed for each clustering method. Each of the cluster categories were compared among the eighteen predictor variables used in the clustering (Figures 4 through 6). The intent here was to identify the unique attributes of each watershed category. A secondary goal was to determine which clustering method was most effective at generating environmentally meaningful differences among categories. Examining these plots, it was apparent that the Bray-Curtis dissimilarity method (Figure 4) was superior to the relative Euclidian distance (Figure 5) and clues methods (Figure 6). The latter two methods resulted in several categories that were poorly differentiated in important attributes, such as population density, imperviousness, and industrial land cover (Figures 5 and 6). Based on these observations, we focus the discussion of watershed categorization results primarily on the Bray-Curtis clustering output.

Description of cluster categories

Cluster analysis produces somewhat arbitrary numerical cluster identifiers (i.e., Cluster 1, 2, etc.), with the key information contained in the attributes of the individual clusters. The eight Bray-Curtis cluster categories vary considerably in their watershed attributes (Table 1). This variation pertains largely to watershed size, spatial location, population density, and land cover (Figure 4, Figure 7).

Clusters 1, 2, and 3 are similar to each other in all having relatively high residential, commercial, and industrial land cover and consequently, high surface imperviousness. Combined, these clusters include 119 watersheds, and could therefore be described as typical watersheds for the study area. These clusters generally include densely populated, low-lying areas that drain into South Bay and Central Bay (Figure 7).

The 41 watersheds in Cluster 1 average 42% residential, 23% industrial, and 13% historic industrial land cover. Cluster 1 has the second highest industrial and historic industry land cover among all clusters, and is also high in impervious surfaces and historic railroads. It includes the previously monitored industrial locations of Zone 4 Line A, and the Ettie Street Pump Station (both in Alameda County) and Richmond Inner Harbor (Contra Costa County). Other representative watersheds in this Cluster include Calabezas Creek (Santa Clara County), and Burlingame Creek (San Mateo County) Sampling in these watersheds could be anticipated to indicate runoff patterns typical of small, relatively urbanized watersheds.

The primary distinction between Clusters 1 and 2 are that Cluster 2 has higher residential land cover (57%), and Cluster 2 has one to four PG&E substations, whereas Cluster 1 watersheds all lack PG&E substations. Thus, comparing watersheds among these clusters might distinguish among potential contaminant loads associated specifically with the presence of PG&E substations. The other difference between 1 and 2 is that watersheds in 2 are often larger in area than 1, though this is not consistently the case (Figures 4 and 7). Among the 43 watersheds in Cluster 2 are Meeker Slough (Contra Costa County), Cordonices Creek (Alameda County), Sunnyvale East (Santa Clara County), and San Bruno Creek (San Mateo County).

The 35 watersheds in Cluster 3 average 45% residential, 13% industrial, and 5% historic industrial land cover. Compared to Clusters 1 and 2, Cluster 3 has lower historic industrial land cover, corresponding to a general absence of historic railroads or water pumping stations in Cluster 3. The other difference is that watersheds in Cluster 3 are generally smaller than in 1 or 2, and often contain Baylands, exposed areas, and sloughs or lagoons. Examples include Point Isabel (Alameda County), Moffett West (Santa Clara County), Foster City Lagoon (San Mateo County), and several unnamed watersheds in Contra Costa County. If the goal were to target monitoring towards historic pollutants due to industrial sources, it would appear that Cluster 1 may be more suitable than Cluster 3, due to the higher density of historic industry and railroads.

Clusters 4 through 8 were much more distinct than Clusters 1, 2, and 3, each characterized by a fairly unique combination of land cover and other attributes. Eleven small to very small watersheds comprise Cluster 4, and with one exception, these are all on or abutting San Francisco International Airport (San Mateo County). These watersheds are all characterized by very high current (68%) and historic (73%) industrial land cover. They consequently have the highest imperviousness among all watersheds, and also have relatively low population density. If any of these watersheds were to contain an untreated and accessible discharge point to the Bay, it would be an interesting candidate for monitoring and characterizing industrial sources.

Clusters 5 and 6 are similar to each other in having high open land cover (61% and 63% open, respectively), and consequently low imperviousness (Figure 4). Cluster 6 contains 22 watersheds, which generally comprise the largest watersheds in the study area (Figure 4, Figure 7). Watersheds in Cluster 6 generally extend to the upland areas, such as the East Bay hills. Examples include the heavily sampled Guadalupe River (Santa Clara

County), San Francisquito Creek (San Mateo County), San Lorenzo Creek (Alameda County), and Wildcat Creek (Contra Costa County). These watersheds have low to moderate residential development (averaging 24%), low imperviousness, and low residential land cover. Cluster 6 watersheds contain a high density of PG&E substations (only exceeded by Cluster 2) and moderate to high modern and historic railroad cover. Given their large area, Cluster 6 watersheds are expected to contain high spatial variability in land cover composition, with higher urban density and impacts in the low-lying areas.

Cluster 5 is distinct from Cluster 6, because Cluster 5 watersheds have very low residential (11%) or commercial (1%) development, and consequently the lowest population density of all watersheds (Figure 4). These watersheds also contain no historic or modern railroads. The 11 watersheds in Cluster 5 include three watersheds that are above reservoirs (Alameda Creek, Guadalupe River, and Coyote Creek, above the respective reservoirs), that do not directly drain into the Bay, and are therefore inappropriate for sampling. The remaining eight watersheds in Cluster 5 are small, nearshore areas, generally comprising reclaimed Baylands and other open spaces. Examples include Bay Farm Island (referred to as AC_unk23), the Palo Alto Golf Course, and Bayfront Park (adjacent to Atherton Creek). It appears that the watersheds in Cluster 5 would be generally inappropriate for stormwater sampling at the Bay margin.

Cluster 7 is the only group containing notable agricultural land cover (43% on average). All but one of these 17 watersheds are in Contra Costa County and drain to Suisun Bay. They would be candidates for monitoring if legacy agricultural sources were of interest. Notable watersheds in Cluster 7 include Mallard Slough, E and W Antioch, Walnut Creek, and Crandall Creek/Zone 5 Line P. Creeks in some of these watersheds have been observed to have elevated pyrethroid pesticides and impacts to local benthic fauna (Amweg et al. 2006).

Cluster 8 is also unique, consisting of just five tiny unnamed watersheds alongshore of the Carquinez Strait. These watersheds average 97% open land cover, and likely represent undeveloped parklands with no current local pollutant sources. One of these watersheds could be a candidate for monitoring as a control site indicative of sediment or pollutant loading due to natural sources and atmospheric deposition.

NMDS results and comparison to cluster analysis

NMDS is an ordination method, and therefore appropriate for describing the underlying variance structure of a data set, and the main variables contributing to differences among samples (in this case watersheds). It is superior to Principal Components Analysis and other parametric methods in having few data structure requirements to result in successful ordinations. We used NMDS for two purposes: 1. to describe which variables best characterize differences among watersheds, and how these variables are related to each other; and 2. to indicate the success of the different clustering methods at generating distinct and compact clusters.

When NMDS ordinations are performed, a calculation is performed of the mismatch between overall distance in the original data set and distance in the ordination results. This calculation, referred to as “stress,” is minimized to obtain the optimal ordination. The final stress is used as a diagnostic indicator of the overall success of the NMDS at characterizing the underlying variation in the data. Stress is measured on a scale from 0 to 100. Stress values above 20 are generally considered to be poor outcomes, and subject to lower interpretive confidence. In our study, the final NMDS stress results were just above 22. Hence, the NMDS results we present were interpreted with caution. However, as NMDS is primarily used as a descriptive technique, we decided to proceed in describing the findings.

Figures 8, 9, and 10 depict the results of the NMDS ordination, in comparison to the three clustering outcomes. The arrows on these figures indicate the direction and relative strength of selected variables in the ordination. We see from looking in the arrows on the x-axis, that the watersheds may generally be distinguished on this axis based on a gradient from industrial land cover with high imperviousness (towards the left) to open land cover. Bray-Curtis Clusters 4 and 8 fit this pattern, as they were previously described as the “outlier” watersheds in these attributes, having almost exclusively industrial and open cover, respectively.

The y-axis appears to correspond to specific developments within the watershed, such as historic railroad and PG&E facilities. Here we see the primary difference between Cluster 3 and Clusters 1 and 2 – though all three of these clusters contain moderate to high industrial land cover, watersheds in Cluster 3 do not contain railroads or PG&E facilities. We might also speculate that some of the watersheds in Cluster 1 constitute the watersheds with greatest overall potential to have legacy contaminant hotspots, associated with historic human activity. This would be based on the observation of Cluster 1 watersheds being located in the direction of industrial activity, impervious surfaces, PG&E facilities, and historic railroads.

Examining Figures 8 through 10, we see a generally weak correspondence among the different clustering methods in characterizing the gradient described in the NMDS. This is consistent with the weak results of the relative Euclidian and clues clustering (Figures 5 and 6), as well as the high stress of the NMDS. The clues method did a particularly poor job, as evident from the broad spread and overlapping pattern of Clusters 1, 5, and 6 in this output (Figure 10). In contrast, the Bray-Curtis output performed relatively well in generating fairly compact and distinct clusters (Figure 8). This finding supports the prior decision to focus on the Bray-Curtis results for characterizing these watersheds (Table 1).

There were some clusters that were consistent among methods. In particular, Clusters 3, 4, and 7 in the Bray-Curtis output corresponded fairly well to clusters 5, 7, and 4 in the Euclidian distance method (Figures 8 and 9). This finding suggests that the small urbanized watersheds (BC Cluster 3), agricultural watersheds (BC Cluster 7), and the industrial watersheds near SF Airport (BC Cluster 4) are clearly distinct from the other watersheds in the study area.

Bivariate approaches to watershed classification

The strength of a clustering approach is its ability to incorporate the information from the multiple attributes that vary across the data set (i.e., the multiple arrows in Figures 8 through 10). The NMDS and cluster analysis results suggest that a complete partitioning of this data set depends on information contained in multiple attributes of the data set. For example, clusters 1 through 3 all contain high residential land cover and imperviousness but are differentiated based on railroads, historic industry, and PG&E substations. In contrast, clusters 6 and 7 are characterized by high open and agricultural land cover (Table 1).

Nevertheless, if management considerations dictate that classification should be based on only one or two of the attributes, this may be readily accomplished with the data set assembled. For example, a bivariate approach would generate a set of categories based on two of the available attributes. To illustrate the concept, we partitioned the data set according to a combination of percent imperviousness and one other attribute. Percent imperviousness was selected because of its importance for watershed contaminant loading, and because it explains a relatively high proportion of the variation in the data set. This relatively high importance is illustrated in the long arrow for this attribute in Figures 8 through 10.

To most effectively partition the variance in the data set, the second variable should explain substantial additional variation beyond that explained by percent imperviousness. This is apparent in the NMDS plots as arrows perpendicular to the percent imperviousness arrow. Historic railroads and PG&E facilities both fit this criteria, being largely vertical, whereas percent imperviousness is horizontal (Figure 8).

We generated a six category classification based on percent imperviousness and PG&E facilities (Figure 11). Three category divisions were made for percent imperviousness (A: < 30%; B: >=30% to < 50%; and C: >=50%). Within each of these categories, two subdivisions were generated based on presence (1) or absence (0) of PG&E facilities. We generated another six category classification based on percent imperviousness and historic railroads (Figure 12). For this scheme, the three percent imperviousness categories described above (i.e., A, B, and C) were each divided into two subcategories based on presence (1) or absence (0) of any historic railroads within the watershed.

The resulting bivariate classifications (Figures 11 and 12) corresponded to some extent to the cluster analysis map (Figure 7). In particular, many of the large watersheds in clusters 5, 6, and 8 also fell into the low imperviousness category. However, some of the information in the cluster analysis was missing from the bivariate results. For example, percent agricultural land cover clearly differentiated cluster 7 from the other clusters, and this attribute would appear important for certain kinds of contaminants. Nevertheless, this information was not apparent from the bivariate classifications.

Table 1. Description of eight watershed clusters generated using Bray-Curtis distance with Ward's linkage method.

Cluster #	Number of watersheds	Description
1	41	High commercial and residential land cover and imperviousness. High historic industry and railroads. No PG&E facilities. Moderate area.
2	43	High commercial and residential land cover and imperviousness. High historic industry and railroads. One to four PG&E facilities. Large area.
3	35	High commercial and residential land cover and imperviousness. Low historic industry or railroads. Smaller area.
4	11	Small, sparsely populated, predominantly industrial, highest historic industrial and imperviousness. Located around San Francisco Airport and Brisbane.
5	11	Sparsely populated, low development, high open land cover, no railroads, "green space." Located adjacent to Bay or in undeveloped uplands.
6	22	Largest watersheds, with moderate population density, high open land cover, and low imperviousness.
7	17	High agricultural land cover, lower rainfall, draining to Carquinez Strait and Suisun Bay.
8	5	Small, sparsely populated, predominantly open, containing historic railroad, and draining to Carquinez Strait.

Figure Captions

Figure 1. Map of the study area.

Figure 2. Dendogram of cluster analysis using Bray-Curtis dissimilarity measure, with rectangles indicating the 8 clusters selected.

Figure 3. Dendogram of cluster analysis using relative Euclidian distance dissimilarity measure, with rectangles indicating the 8 clusters selected.

Figure 4. Box and whiskers plots for the eight clusters selected in the Bray-Curtis dissimilarity cluster analysis. Each plot indicates the results of one of 18 numeric metrics. Metrics were all transformed and rescaled to range from 0 to 1, as described in methods.

Figure 5. Box and whiskers plots for the eight clusters selected in the relative Euclidian dissimilarity cluster analysis.

Figure 6. Box and whiskers plots for the seven clusters selected in the clues algorithm cluster analysis.

Figure 7. Map of the study area, indicating the watershed categorization among clusters. Results are for the Bray-Curtis dissimilarity cluster analysis (Figures 1 and 3), with color coding indicating which cluster each watersheds falls into.

Figure 8. NMDS ordination results with Bray Curtis clustering outcomes indicated by symbols. Arrows indicate direction and relative magnitude of labeled variables. Variables listed with arrows in Figures 7 to 9 were chosen because they were strongly associated with clustering or to illustrate specific gradients (e.g., agricultural land cover, PG&E facility).

Figure 9. NMDS ordination results with Relative Euclidian distance clustering outcomes indicated by symbols. Arrows indicate direction and relative magnitude of labeled variables.

Figure 10. NMDS ordination results with Clues algorithm clustering outcomes indicated by symbols. Arrows indicate direction and relative magnitude of labeled variables.

Figure 11. Bivariate classification results based on percent imperviousness and PG&E facilities.

Figure 12. Bivariate classification results based on percent imperviousness and presence of historic railroads.

References

Ackerman, J. T., J. Y. Takekawa, C. A. Eagles-Smith, and S. A. Iverson. 2008. Mercury contamination and effects on survival of American avocet and black-necked stilt chicks in San Francisco Bay. *Ecotoxicology* **17**:103-116.

Amweg, E. L., D. P. Weston, J. You, and M. J. Lydy. 2006. Pyrethroid insecticides and sediment toxicity in urban creeks from California and Tennessee *Environ. Sci. Technol.* **40**:1700 -1706.

Anderson, D. W., J.R. Jehl, R.W. Risebrough, L.A. Woods, L.R. Deweese, and W. G. Edgecombe. 1975. Brown pelicans: improved reproduction off the southern California coast. *Science* **190**:806-808.

Association of Bay Area Governmants (ABAG), 1997. Regional Existing Land Use in 1995: Data for Bay Area Counties. http://www.abag.ca.gov/

Brar, N. K., C. Waggoner, J. A. Reyes, R. Fairey, and K. M. Kelley. 2010. Evidence for thyroid endocrine disruption in wild fish in San Francisco Bay, California, USA. Relationships to contaminant exposures. *Aquat. Toxicol.* **96**:203-215.

Bridges, T. S., W. J. Berry, S. D. Sala, P. B. Dorn, S. J. Ells, T. H. Gries, D. S. Ireland, E. M. Maher, C. A. Menzie, L. M. Porebski, and J. Stronkhorst. 2005. A framework for assessing and managing risks from contaminated sediments. Pages 227-266 *in* R. J. Wenning, G. E. Batley, C. G. Ingersoll, and D. W. Moore, editors. Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments. SETAC, Pensacola, FL.

Brown, F. R., J. Winkler, P. Visita, J. Dhaliwal, and M. Petreas. 2006. Levels of PBDEs, PCDDs, PCDFs, and coplanar PCBs in edible fish from California coastal waters. *Chemosphere* **64**:276-286.

Bulley, H. N. N., J. W. Merchant, D. B. Marx, J. C. Holz, and A. A. Holz. 2007. A gis-based approach to watershed classification for Nebraska reservoirs. *Journal of the American Water Resources Association* **43**:605-621.

Chang, F., W. Qiu, R. H. Zamar, R. Lazarus, and X. Wang. 2010. clues: An R Package for Nonparametric Clustering Based on Local Shrinking. *Journal of Statistical Software* **33, Issue 4, Feb 2010**:15 pp.

City of San Jose and EOA, Inc. July 2003. Year Two Case Study Investigating Elevated Levels of PCBs in Storm Drain Sediments in San Jose, California. Prepared for: Santa Clara Valley Urban Runoff Pollution Prevention Program. pp 32.

Collier, T. K., L. L. Johnson, C. M. Stehr, M. S. Myers, and J. E. Stein. 1998. A comprehensive assessment of the impacts of contaminants on fish from an urban waterway. *Mar. Environ. Res.* **46**:243-247.

Covelli, S., J. Faganeli, M. Horvat, and A. Brambati. 2001. Mercury contamination of coastal sediments as the result of long-term cinnabar mining activity (Gulf of Trieste, northern Adriatic sea). *Appl Geochem* **16**:541-558.

Cox, T. F., and M. A. A. Cox. 1994, 2001. Multidimensional Scaling. Chapman & Hall.

David, N., L. J. McKee, F. J. Black, A. R. Flegal, C. H. Conaway, D. H. Schoellhamer, and N. K. Ganju. 2009. Mercury concentrations and loads in a large river system tributary to San Francisco Bay, California, USA. *Environ. Toxicol. Chem.* **In press**.

Davis, J. A., M. D. May, B. K. Greenfield, R. Fairey, C. Roberts, G. Ichikawa, M. S. Stoelting, J. S. Becker, and R. S. Tjeerdema. 2002. Contaminant concentrations in sport fish from San Francisco Bay, 1997. *Mar Pollut Bull* **44**:1117-1129.

Eilers, J. M., G. E. Glass, K. E. Webster, and J. A. Rogalla. 1983. Hydrologic control of lake susceptibility to acidification. *Can. J. Fish. Aquat. Sci.* **40**:1896-1904.

Eisler, R. 1987. Polycyclic aromatic hydrocarbon hazards to fish, wildlife, and invertebrates: a synoptic review. Biological Report 85 (1. 11), US Fish and Wildlife Service, Laurel, Maryland.

EOA, Inc. 2004. Case Study Investigating PCBs in Storm Drain Sediments from Colma Creek, Colma, California. Prepared for: San Mateo Countywide Stormwater Pollution Prevention Program. May 2004. pp 16.

EOA, Inc. 2007. Summary of Polychlorinated Biphenyls (PCBs) Data in Sediment Collected from Richmond, California Streets and Storm Drains. Prepared for: City of Richmond. October 2007. pp. 25.

EOA, Inc. April 15, 2002. Case Study Investigating Elevated Levels of PCBs in Storm Drain Sediments in San Mateo County. Prepared for San Mateo Countywide Stormwater Pollution Program. pp 31.

Gergel, S. E., E. M. Bennett, B. K. Greenfield, S. King, C. A. Overdevest, and B. Stumborg. 2004. A test of the environmental kuznets curve using long-term watershed inputs. *Ecological Applications* **14**:555-570.

Greenfield, B. K., C. H. Conaway, A. Jahn, M. C. Marvin-DiPasquale, C. A. Eagles-Smith, J. T. Ackerman, D. P. Krabbenhoft, and M. B. Sandheinrich. 2009. Spatial and temporal patterns of mercury contamination in San Francisco Bay sediment and biota. Regional Monitoring Program, Oakland, CA.

Greenfield, B. K., and A. Jahn. 2010. Mercury in San Francisco Bay forage fish. *Environ Pollut* **In press**.

Gunther, A., Salop, P., Bell, D., Feng, A., Wiegel, J., and Wood, R. 2001. Initial Characterization of PCB, Mercury, and PAH Contamination in the Drainages of Western, Alameda County, CA. Prepared for: Alameda Countywide Clean Water Program. March 2001. pp 45.

Haycock, N. E., G. Pinay, and C. Walker. 1993. Nitrogen-retention in river corridors - European perspective. *Ambio* **22**:340-346.

Hornberger, M. I., S. N. Luoma, A. van Geen, C. Fuller, and R. Anima. 1999. Historical trends of metals in the sediments of San Francisco Bay, California. *Mar Chem* **64**:39-55.

Kannan, K., H. Nakata, R. Stafford, G. R. Masson, S. Tanabe, and J. P. Giesy. 1998. Bioaccumulation and Toxic Potential of Extremely Hydrophobic Polychlorinated Biphenyl Congeners in Biota Collected at a Superfund Site Contaminated with Aroclor 1268 *Environ. Sci. Technol.* **32**:1214-1221.

Kennish, M. J., and B. E. Ruppel. 1998. Organochlorine contamination in selected estuarine and coastal marine finfish and shellfish of New Jersey. *Water Air Soil Pollut.* **101**:123-136.

KLI and EOA, 2002. Joint Stormwater Agency Project to Study Urban Sources of Mercury, PCBs and Organochlorine Pesticides. Prepared for: Santa Clara Valley Urban Runoff Pollution Prevention Program, Contra Costa Clean Water Program, San Mateo Countywide Stormwater Pollution Prevention Program, Marin County Stormwater Pollution Prevention Program, Vallejo.

Kleinfelder Inc., 2005. Sediment Sampling Report: Ettie Street Pump Station Watershed, Oakland, California. Prepared for: City of Oakland PWA -ESD. July 2005. pp 123.

Kleinfelder Inc., 2006. Private Property Sediment Sampling Report: Ettie Street Watershed, Oakland, California. Prepared for: City of Oakland PWA-WSD. January 2006. pp 59.

Kuzyk, Z. A., P. V. Hodson, S. M. Solomon, and K. J. Reimer. 2005a. Biological responses to PCB exposure in shorthorn sculpin from Saglek Bay, Labrador. *Sci. Total Environ.* **351**:285-300.

Kuzyk, Z. A., J. P. Stow, N. M. Burgess, S. M. Solomon, and K. J. Reimer. 2005b. PCBs in sediments and the coastal food web near a local contaminant source in Saglek Bay, Labrador. *Sci. Total Environ.* **351**:264-284.

Lauenstein, G. G., and K. D. Daskalakis. 1998. U.S. long-term coastal contaminant temporal trends determined from mollusk monitoring programs, 1965-1993. *Mar Pollut Bull* **37**:6-13.

Lewicki, M., and L. J. McKee. 2009. Watershed specific and regional scale suspended sediment loads for Bay Area small tributaries. A technical report for the Sources Pathways and Loading Workgroup of the Regional Monitoring Program for Water Quality. SFEI Contribution #566, San Francisco Estuary Institute, Oakland, CA.

Linkov, I., D. Burmistrov, J. Cura, and T. S. Bridges. 2002. Risk-based management of contaminated sediments: consideration of spatial and temporal patterns in exposure modeling. *Environ. Sci. Technol.* **36**:238-246.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MJM Software Design, Gleneden Beach, OR. www.pcord.com

McKee, L., J. Leatherbarrow, and J. Oram. 2005. Concentrations and loads of mercury, PCBs, and OC pesticides in the lower Guadalupe River, San Jose, California: Water Years 2003 and 2004. SFEI Contribution Number 409, SFEI, Oakland, CA. http://www.sfei.org/watersheds/reports/409_GuadalupeRiverLoadsYear2.pdf

McKee, L., P. Mangarella, B. Williamson, J. Hayworth, and L. Austin. 2006. Review of Methods to Reduce Urban Stormwater Loads: Task 3.4. A Technical Report of the Regional Watershed Program. SFEI Contribution #429, San Francisco Estuary Institute, Oakland, CA.

Minchin, P. R. 1987. An evaluation of relative robustness of techniques for ecological ordinations. *Vegetatio* **69**:89-107.

National Land Cover Dataset (NLCD) 2001. Impervious Surface: Percent Imperviousness. http://gisdata.usgs.gov/website/mrlc/viewer.htm

OEHHA. 1997. Health advisory on catching and eating fish: interim sport fish advisory for San Francisco Bay. *in*. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://www.oehha.org/fish/nor_cal/int-ha.html

Salop, P., Abu-Saba, K., Gunther, A., and Feng, A., September 2002. 2000-01 Alameda County Watershed Sediment Sampling Program: Two-Year Summary and Analysis. Prepared for: Alameda County Clean Water Program. pp 40.

Salop, P., Hardin, D., Abu-Saba, K., Gunther, A., and Feng, A. 2002. Analysis of 2000-01 Source Investigations in Ettie Street Pump Station and Glen Echo Creek Watersheds, Oakland, California. Prepared for: Alameda Countywide Clean Water Program. August 2002. pp 33.

SFBRWQCB. 2006. Mercury in San Francisco Bay Total Maximum Daily Load (TMDL) Proposed Basin Plan Amendment and Staff Report for Revised Total Maximum Daily Load (TMDL) and Proposed Mercury Water Quality Objectives. *Final Report.* California Regional Water Quality Control Board San Francisco Bay Region, Oakland, CA. http://www.swrcb.ca.gov/rwqcb2/TMDL/sfbaymercurytmdl.htm

SFBRWQCB. 2008. Total Maximum Daily Load for PCBs in San Francisco Bay. Staff Report for Proposed Basin Plan Amendment San Francisco Bay Regional Water Quality Control Board, Oakland, CA. http://www.waterboards.ca.gov/sanfranciscobay/TMDL/SFBayPCBs/pcbs_tmdl_project_report010804.pdf

SFBRWQCB. 2009. California Regional Water Quality Control Board San Francisco Bay Region Municipal Regional Stormwater NPDES Permit, Order R2-2009-0074, NPDES Permit No. CAS612008. Adopted October 14, 2009. 279 pp. http://www.waterboards.ca.gov/sanfranciscobay/board_decisions/adopted_orders/2009/R2-2009-0074.pdf

SFEI, 2010. Regional Stormwater Monitoring and Urban BMP Evaluation. San Francisco Estuary Institute, Oakland CA.San Mateo Countywide Stormwater Pollution Prevention Program. June 2003. pp 21.

SFEI, 2010. A BMP tool box for reducing Polychlorinated biphenyls (PCBs) and Mercury (Hg) in municipal stormwater. San Francisco Estuary Institute, Oakland CA.

Smith, V. H. 1998. Cultural eutrophication of inland, estuarine, and coastal waters. Pages 7-49 *in* P. a. Groffman, editor. Successes, Limitations, and Frontiers in Ecosystem Science. Springer-Verlag, New York.

State Water Resources Control Board. 2009. California's 2006 Clean Water Act Section 303(d) List of Water Quality Limited Segments. *in*. State Water Resources Control Board, Sacramento, CA. http://www.swrcb.ca.gov/water_issues/programs/tmdl/303d_lists2006_epa.shtml

Strom, D. G., and G. A. Graves. 2001. A comparison of mercury in estuarine fish between Florida Bay and the Indian River Lagoon, Florida USA. *Estuaries* **24**:597-609.

Suchanek, T. H., J. Cooke, K. Keller, S. Jorgensen, P. J. Richerson, C. A. Eagles-Smith, E. J. Harner, and D. P. Adam. 2009. A mass balance mercury budget for a mine-dominated lake: Clear Lake, California. *Water Air Soil Pollut.* **196**:51-73.

Tay, K.-L., S. J. Teh, K. Doe, K. Lee, and P. Jackman. 2003. Histopathologic and histochemical biomarker responses of Baltic Clam, *Macoma balthica*, to contaminated Sydney Harbour sediment, Nova Scotia, Canada. *Environ. Health Persp.* **111**:273-280.

Thompson, B., T. Adelsbach, C. Brown, J. Hunt, J. Kuwabara, J. Neale, H. Ohlendorf, S. Schwarzbach, S. R., and K. Taberski. 2007. Biological effects of anthropogenic contaminants in the San Francisco Estuary. *Environ. Res.* **105**:156-174.

Trimble, S. W. 2003. Historical hydrographic and hydrologic changes in the San Diego creek watershed, Newport Bay, California. *Journal Of Historical Geography* **29**:422-444.

van Geen, A., and S. N. Luoma. 1999. The impact of human activities on sediments of San Francisco Bay, California: an overview. *Mar Chem* **64**:1-6.

Young, T. C., and J. L. Stoddard. 1996. The Temporally Integrated Monitoring of Ecosystems (TIME) project design 1. Classification of northeast lakes using a combination of geographic, hydrogeochemical, and multivariate techniques. *Water Resources Research* **32**:2517-2528.

Appendix A – Numbered watershed names and categories in Figures 7, 11, and 12

ID Number	Watershed Name	Figure 7 Category	Figure 11 Category	Figure 12 Category
Fig. 7, 11, 12		Bray Curtis Clusters	Impervious and PG&E	Impervious and railroad
1	ACFC_Zone 4 Line A	1	C.0	C.1
2	ACFC_Zone 5 Line F-1	1	C.0	C.1
3	ACFC_Zone 5 Line J-3 Pump Station	1	C.0	C.0
4	ACFC_Zone 5 Line P and Zone 6 Line N	2	B.1	B.1
5	AC_unk01	3	B.0	B.0
6	AC_unk02	3	B.0	B.0
7	AC_unk03	1	C.0	C.1
8	Emeryville Watershed	1	C.0	C.1
9	AC_unk05	3	B.0	B.0
11	Piedmont Watershed North	3	B.1	B.0
12	AC_unk08	3	C.0	C.0
13	Port of Oakland	2	C.1	C.1
15	Piedmont Watershed South	2	C.1	C.1
16	AC_unk12	4	C.0	C.1
17	AC_unk13	1	C.0	C.1
18	AC_unk14	1	C.0	C.1
19	AC_unk15	1	C.0	C.1
20	AC_unk16	3	B.0	B.0
21	AC_unk17	1	C.0	C.1
22	AC_unk18	3	C.0	C.0
23	AC_unk19	3	B.0	B.0
24	AC_unk20	3	C.0	C.0
25	AC_unk22	3	C.0	C.0
26	AC_unk24	1	C.0	C.1
27	AC_unk25	5	A.0	A.0
29	AC_unk27	5	C.0	C.0
30	AC_unk28	2	C.1	C.1
36	Adobe Creek	6	A.1	A.1
37	Agua Fria and Torogas Creek and Scott Creek	6	A.1	A.1
38	Arroyo Viejo	2	A.1	A.1
39	Atherton Creek	2	B.1	B.1
40	Barron Creek	1	B.0	B.1
41	Baxter Creek	2	C.1	C.1
42	Bayfront Park	5	A.0	A.0
43	Belmont Creek	2	B.1	B.1
44	Belmont Slough	3	B.0	B.0
45	Blackberry and Marin Creeks_A	2	B.1	B.1
46	Bockman Canal	2	C.1	C.1
47	Borel Creek	2	B.1	B.1
48	Burlingame Creek	1	A.0	A.1
49	Cerrito Creek	2	B.1	B.0
50	Coast Casey Forebay	3	C.0	C.1
51	Codornices Creek	2	A.1	A.1
52	Colma Creek	2	B.1	B.1
53	Cordilleras Creek	2	A.1	A.1

ID Number Fig. 7, 11, 12	Watershed Name	Figure 7 Category Bray Curtis Clusters	Figure 11 Category Impervious and PG&E	Figure 12 Category Impervious and railroad
54	Crandall Creek and ACFC_Zone 5 Line P	7	B.0	B.0
56	Derby and Potter Creeks_A	2	C.1	C.1
57	Easton Creek	2	B.1	B.1
58	Elmhurst Creek_A	2	C.1	C.1
59	Estudillo Canal	2	C.1	C.1
60	Ettie Street Pump Station_A	1	C.0	C.1
61	Garrity Creek	1	B.0	B.1
62	Glen Echo Creek	3	B.0	B.0
63	Green Hills Creek	2	B.1	B.1
64	Guadalupe River	6	B.1	B.1
65	Guadalupe Valley Creek	1	A.0	A.1
66	Herman Slough and Castro Creek	1	C.0	C.1
67	Hoffman Channel	2	B.1	B.0
68	Agua Caliente	6	A.1	A.1
69	Laurel Creek	1	B.0	B.1
70	Leslie Creek	1	C.0	C.1
71	Lion Creek	1	B.0	B.1
72	Coyote Creek Above Anderson Dam	5	A.1	A.0
73	Lower Sulphur Creek	1	C.0	C.1
74	Mallard Slough	7	A.0	A.0
75	Marina Lagoon	3	C.0	C.0
76	Matadero Creek	1	A.0	A.1
77	Meeker Slough	2	C.1	C.1
78	Millbrae Creek	1	C.0	C.1
79	Mills Creek	1	B.0	B.1
80	Treasure Island	3	C.0	C.0
81	West Antioch	7	A.1	A.0
83	Moffett West	3	B.0	B.1
85	Oyster Point	4	C.0	C.0
86	Palo Alto Golf Course	5	B.0	B.0
87	Peralta and Courtland and Seminary Creeks	2	C.1	C.1
88	Permanente Creek	6	A.1	A.1
89	Pinole Creek	6	A.0	A.1
90	Pinole Shores	1	B.0	B.1
91	Point Isabel	3	C.0	C.0
92	Point Richmond	1	B.0	B.1
93	Point San Pablo Peninsula North	1	A.0	A.1
94	Poplar Creek	2	B.1	B.1
95	Pulgas Creek	2	B.1	B.1
96	Redwood Ck and Arroyo Ojo de Agua Ck	2	B.1	B.1
97	Redwood Shores Lagoon Water	3	B.0	B.0
98	Refugio Creek	2	A.1	A.1
99	East Antioch	7	B.1	B.0
101	Richmond Inner Harbor	1	C.0	C.1
102	SMC_unk01	2	B.1	B.1
103	SMC_unk02	1	A.0	A.1
105	SMC_unk04	4	C.0	C.0
106	SMC_unk05	4	C.0	C.0

ID Number Fig. 7, 11, 12	Watershed Name	Figure 7 Category Bray Curtis Clusters	Figure 11 Category Impervious and PG&E	Figure 12 Category Impervious and railroad
107	SMC_unk06	4	C.0	C.1
108	SMC_unk07	4	C.0	C.1
109	SMC_unk08	4	C.0	C.1
110	SMC_unk09	4	C.0	C.0
111	SMC_unk10	3	B.0	B.0
112	SMC_unk11	3	C.0	C.0
113	SMC_unk12	1	C.0	C.1
114	SMC_unk14	3	C.0	C.0
115	SMC_unk15	1	C.0	C.1
116	SMC_unk16	2	B.1	B.1
117	SMC_unk17	1	A.0	A.1
118	SMC_unk18	3	C.1	C.1
119	SMC_unk19	1	C.0	C.1
120	San Bruno Creek	2	B.1	B.1
121	San Francisco International Airport A	4	C.0	C.0
122	San Francisco International Airport B	4	C.0	C.0
123	San Francisquito Creek	6	A.1	A.1
124	San Leandro Creek Above Lake Chabot	6	A.1	A.0
125	San Lorenzo Creek	6	A.1	A.1
126	San Mateo Creek Above Reservoir	6	A.1	A.0
127	San Pablo Creek Above Reservoir	6	A.1	A.0
128	San Tomas	1	B.0	B.1
129	Sanchez Creek	2	A.1	A.1
130	Sanjon de los Alisos A	2	C.1	C.1
131	Santa Fe Channel	2	C.1	C.1
132	Sausal Creek	2	A.1	A.1
133	Schoolhouse Creek	1	B.0	B.1
134	Seal Slough	3	C.0	C.0
135	Sewage Treatment Plant	5	A.0	A.0
136	Walnut Creek	7	A.1	A.0
137	Stevens Creek	6	A.1	A.1
138	Strawberry Creek	2	B.1	B.0
140	Temescal Creek	2	A.1	A.1
141	Unknown_240	3	B.0	B.0
142	Unknown_241	3	B.0	B.0
143	Unknown_244	3	B.0	B.0
144	Unknown_245	3	A.0	A.0
145	Unknown_246	3	B.0	B.0
146	Unknown_247	3	B.0	B.0
147	Unknown_248	5	A.0	A.0
148	Unknown_251	8	A.0	A.1
151	Alhambra Creek	6	A.1	A.1
152	Unknown_256	7	A.0	A.0
153	Unknown_257	7	A.0	A.0
154	Unknown_258	8	A.0	A.1
155	Unknown_259	7	A.0	A.0
156	Unknown_260	7	A.0	A.0
157	Unknown_261	3	B.0	B.0

ID Number Fig. 7, 11, 12	Watershed Name	Figure 7 Category Bray Curtis Clusters	Figure 11 Category Impervious and PG&E	Figure 12 Category Impervious and railroad
158	Unknown_262	7	A.0	A.0
159	Unknown_263	2	B.1	B.1
160	Unknown_264	8	A.0	A.1
161	Unknown_265	3	B.1	B.0
162	Unknown_266	7	B.0	B.0
163	Unknown_267	7	A.0	A.0
164	Unknown_268	7	B.1	B.0
165	Unknown_272	7	A.0	A.0
166	Unknown_273	7	A.0	A.0
167	Unknown_274	7	A.0	A.0
168	Unknown_275	7	A.0	A.0
174	Unknown_284	3	A.0	A.0
175	Visitacion Point	1	C.0	C.1
176	Old Alameda Creek	2	B.1	B.1
177	Yerba Buena Island	3	A.0	A.0
178	Mount Diablo Creek	6	A.0	A.0
179	Kirker Creek	6	A.0	A.0
180	Unknown_271	2	C.1	C.0
181	Alameda Creek Above Reservoir	5	A.0	A.0
182	Point San Pablo Peninsula West	1	A.0	A.1
183	AC_unk23	5	B.0	B.0
184	AC_unk21	5	A.0	A.0
185	SMC_unk13	3	A.0	A.0
186	Foster City Lagoon Water	3	C.0	C.0
187	Rodeo Creek	6	A.1	A.1
188	Canada del Cierbo	6	A.0	A.1
189	Wildcat Creek	6	A.0	A.1
190	Rheem Creek	1	B.0	B.1
191	Point Pinole	2	A.1	A.1
203	San Leandro Creek Below Lake Chabot	2	C.1	C.1
294	Guadalupe River Above Reservoir	5	A.0	A.0
295	Lower Coyote Creek (below Dam)	6	A.1	A.1
296	San Mateo Creek	2	A.1	A.1
297	Alameda Creek	6	A.1	A.1
299	San Pablo Creek	6	A.1	A.1
341	Sunnyvale West	1	C.0	C.1
342	Sunnyvale East	2	C.1	C.1
343	Calabazas Creek	1	B.0	B.1
345	Refugio North	1	A.0	A.1
349	Davis Point	1	A.0	A.1
None	Unknown_253MergeManual	8	A.0	A.1
None	Unknown_252MergeManual	8	A.0	A.1
None	Unknown_278MergeManual	6	A.1	A.1
None	SMC_unk03MergeManual	4	C.0	C.1

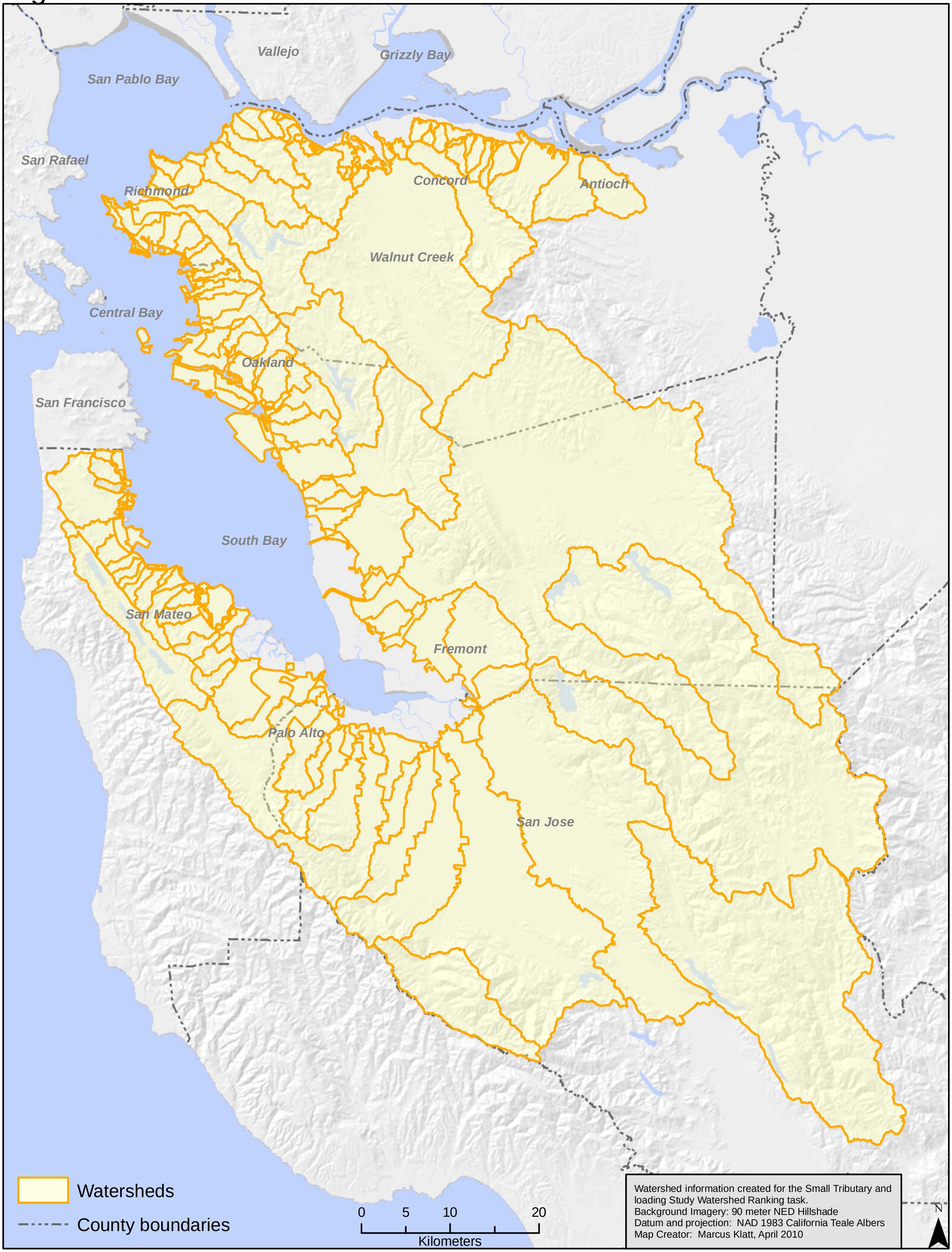
Appendix B – Background Information

Coastal ecosystems around the world are the focus of urbanization, industrialization, agriculture, transport (rail, road, and shipping) and waste disposal and as such are subject to loads of suspended sediments, nutrients, pathogens, and trace organic and metallic wastes (Haycock et al. 1993, Lauenstein and Daskalakis 1998, Smith 1998, Covelli et al. 2001, Linkov et al. 2002, Trimble 2003, Bridges et al. 2005, Kuzyk et al. 2005b). In the recent century, the advancement of chemical process technology lead to the use and synthesis of a number of persistent metals including mercury (Hg), copper, lead, zinc and silver and organic compounds including polyaromatic hydrocarbons (PAHs), organochlorine (e.g. DDT), organophosphate (e.g. melathion), and synthetic pyrethroid pesticides (e.g.), polychlorinated byphenyls (PCBs), and polybrominated diphenylethers (PBDEs). Although these substances have provided for many useful lifestyle improvements, they have also lead to many well documented ecosystem impacts worldwide (Anderson et al. 1975, Eisler 1987, Collier et al. 1998, Kannan et al. 1998, Kennish and Ruppel 1998, Covelli et al. 2001, Strom and Graves 2001, Tay et al. 2003, Gergel et al. 2004, Kuzyk et al. 2005a, Amweg et al. 2006, Brown et al. 2006, Suchanek et al. 2009). San Francisco Bay is one such ecosystem where the balance between technological and economic advancement and ecosystem preservation have resulted in increased concentrations of multiple pollutants (Hornberger et al. 1999, van Geen and Luoma 1999), with effects to local fish and wildlife (Thompson et al. 2007, Ackerman et al. 2008, Brar et al. 2010).

In 2006, in compliance with the Clean Water Act overseen by the U.S. Environmental Protection Agency (U.S. EPA), the state of California included all or some areas of San Francisco Bay[c] on the 303(d) list of water quality limited segments. The current listings are based on organochlorine pesticides (DDT, chlordane, dieldrin), dioxin compounds, exotic species, furan compounds, lead, mercury, nickel, polyaromatic hydrocarbons (PAHs), polychlorinated biphenyl’s (PCBs), sediment toxicity, selenium, and zinc (State Water Resources Control Board 2009). In the 1990s, the California Office of Health Hazard Assessment (OHHA) issued health warnings to those people of catch and consume fish from San Francisco Bay (OEHHA 1997, Davis et al. 2002).

[c] I.e., Central San Francisco Bay and subembayments (e.g., Oakland Inner Harbor, San Leandro Bay, and Central Basin), San Pablo Bay, Suisun Bay, and Carquinez Strait

Figure 1



Vallejo
Grizzly Bay
San Pablo Bay
San Rafael
Richmond
Concord
Antioch
Walnut Creek
Central Bay
Oakland
San Francisco
South Bay
San Mateo
Fremont
Palo Alto
San Jose
Watersheds
County boundaries
0
5
10
20
Kilometers
Watershed information created for the Small Tributary and loading Study Watershed Ranking task.
Background Imagery: 90 meter NED Hillshade
Datum and projection: NAD 1983 California Teale Albers
Map Creator: Marcus Klatt, April 2010
N

Figure 2

Bray Curtis dendogram with no Null and 8 bins

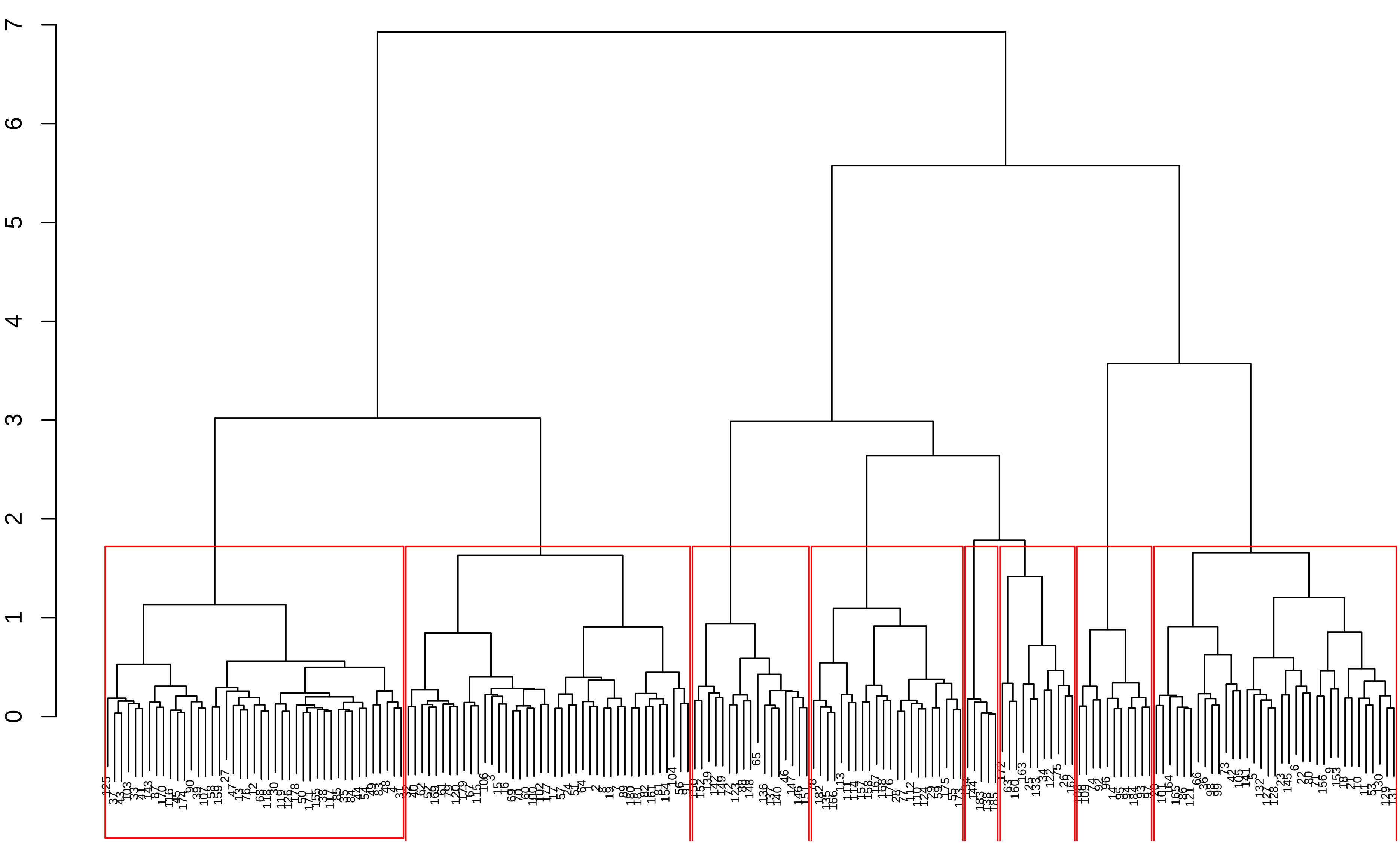
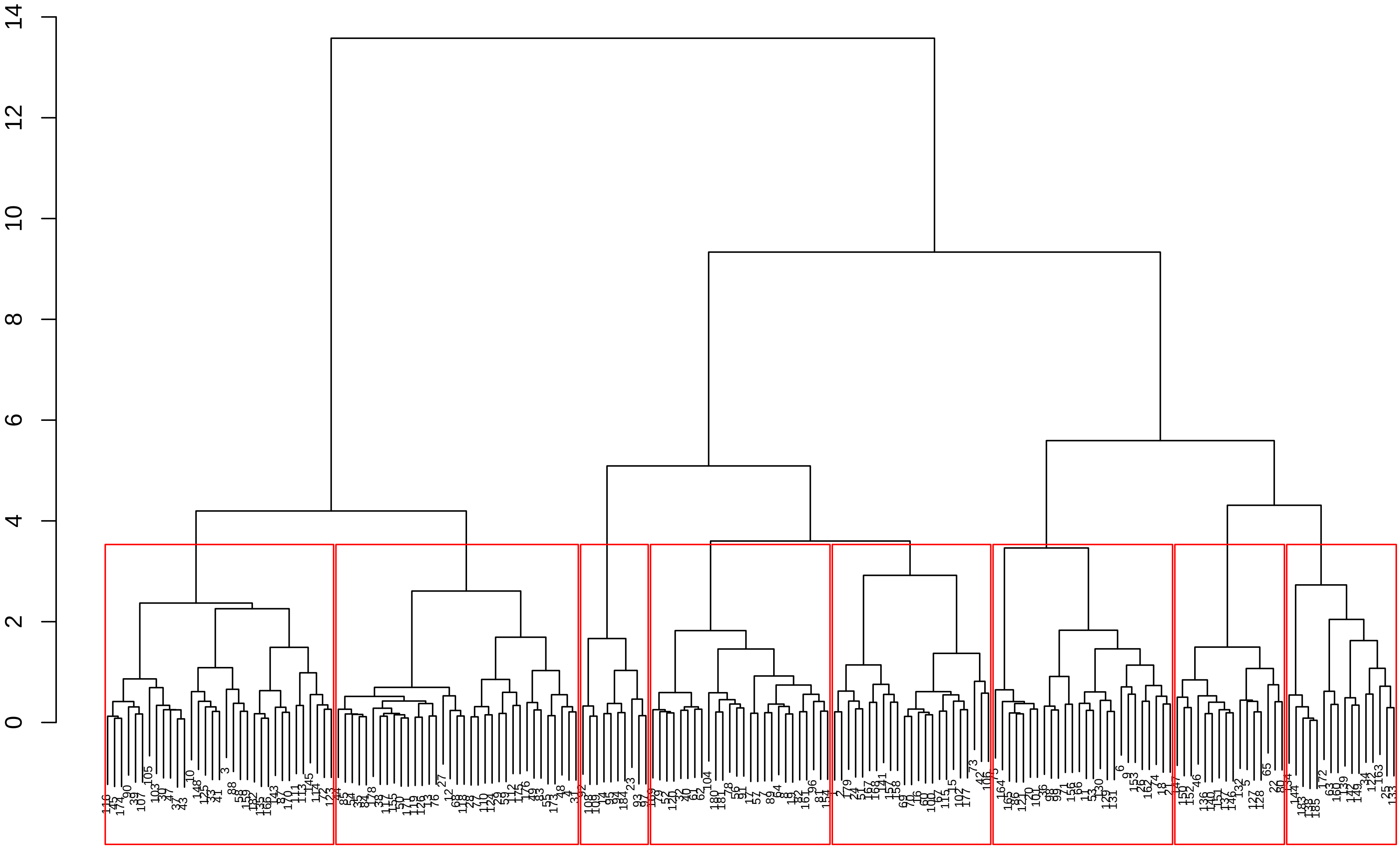


Figure 3

Relative Euclidian Distance dendogram with no Null and 8 bins



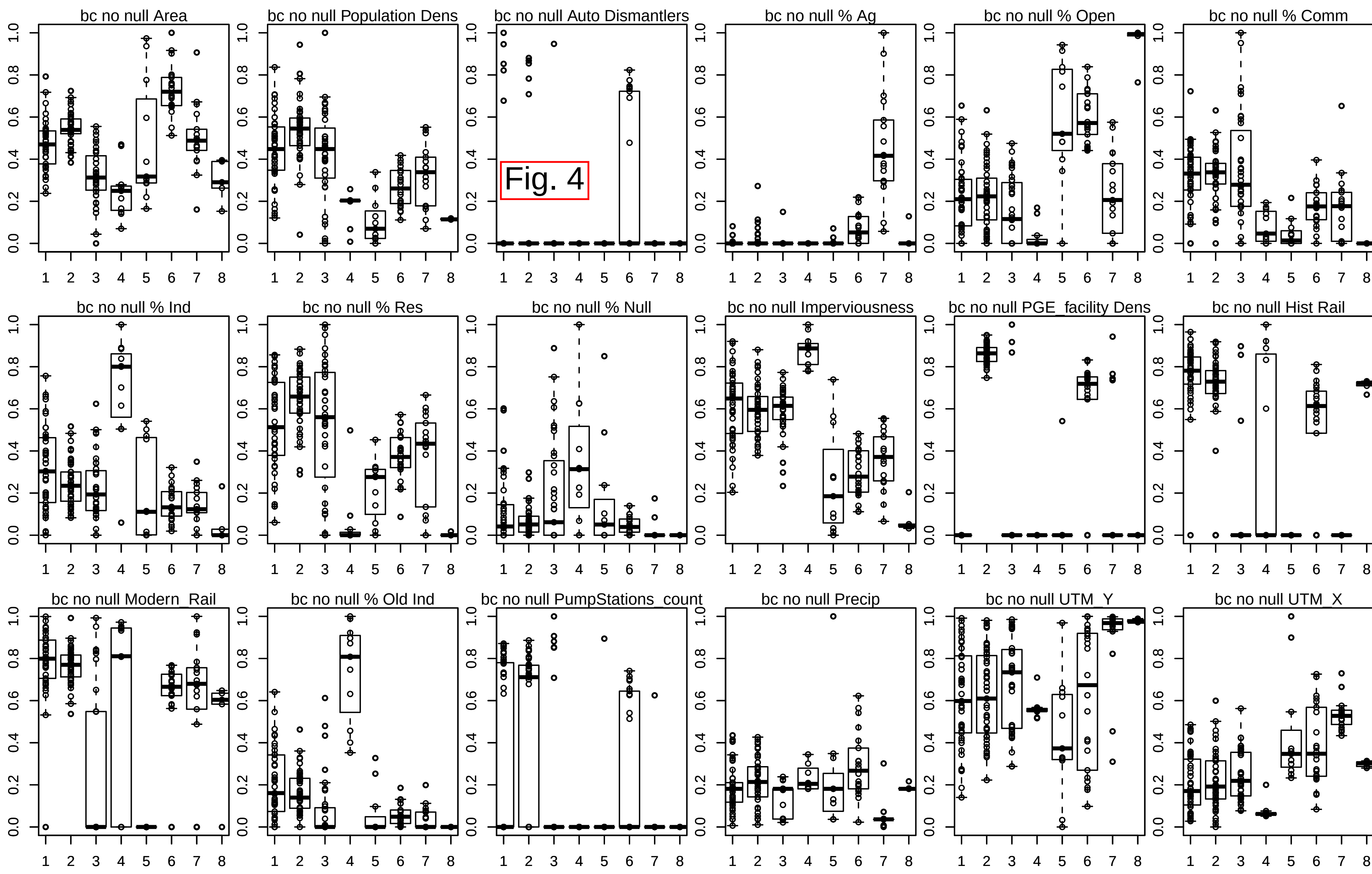
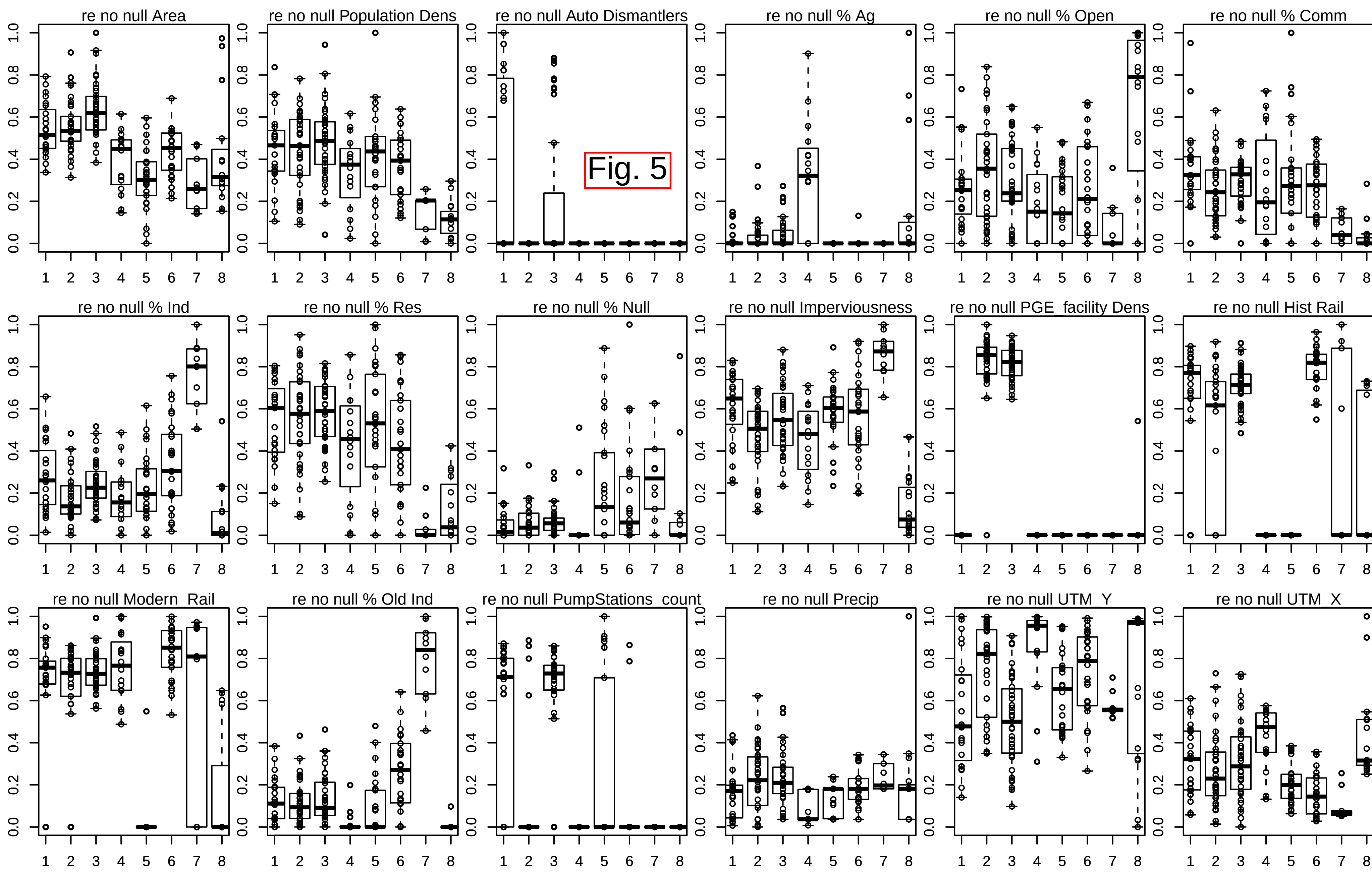

bc no null Area
bc no null Population Dens
bc no null Auto Dismantlers
bc no null % Ag
bc no null % Open
bc no null % Comm
bc no null % Ind
bc no null % Res
bc no null % Null
bc no null Imperviousness
bc no null PGE_facility Dens
bc no null Hist Rail
bc no null Modern_Rail
bc no null % Old Ind
bc no null PumpStations_count
bc no null Precip
bc no null UTM_Y
bc no null UTM_X

Fig. 4


re no null Area
re no null Population Dens
re no null Auto Dismantlers
Fig. 5
re no null % Ag
re no null % Open
re no null % Comm
re no null % Ind
re no null % Res
re no null % Null
re no null Imperviousness
re no null PGE_facility Dens
re no null Hist Rail
re no null Modern_Rail
re no null % Old Ind
re no null PumpStations_count
re no null Precip
re no null UTM_Y
re no null UTM_X

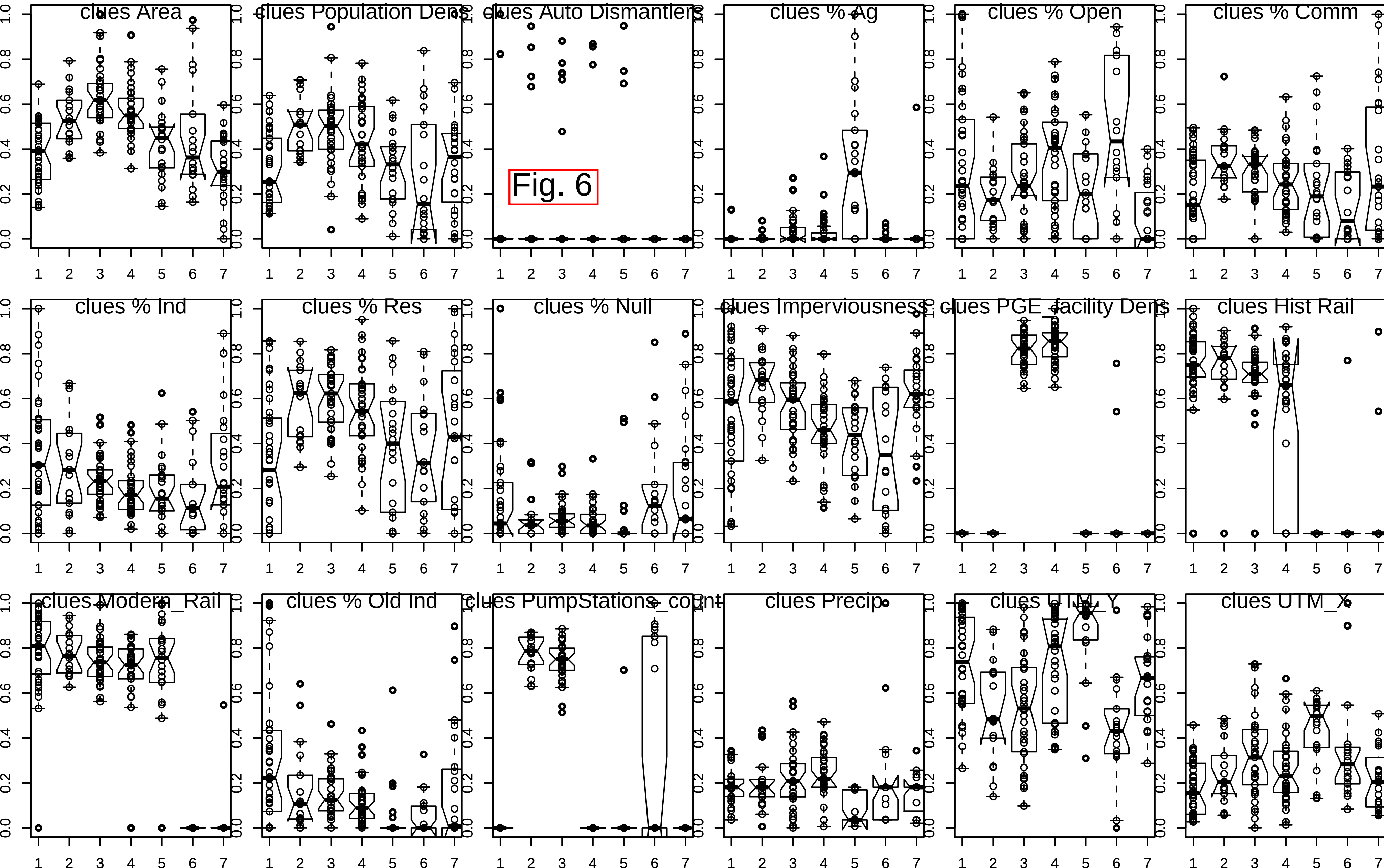

clues Area
clues Population Dens
clues Auto Dismantlers
clues % Ag
clues % Open
clues % Comm
clues % Ind
clues % Res
clues % Null
clues Imperviousness
clues PGE_facility Dens
clues Hist Rail
clues Modern_Rail
clues % Old Ind
clues PumpStations_count
clues Precip
clues UTM_Y
clues UTM_X

Fig. 6

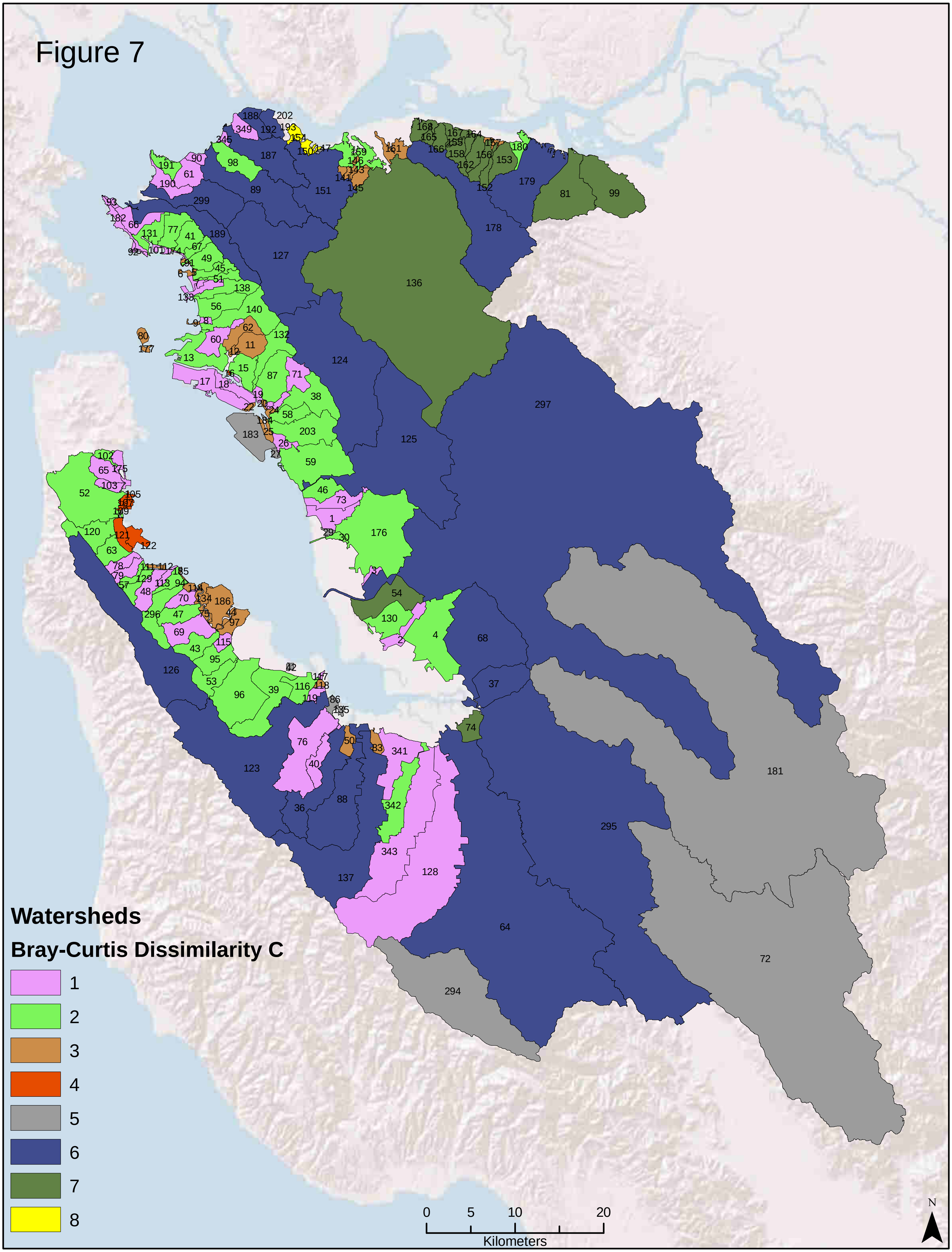
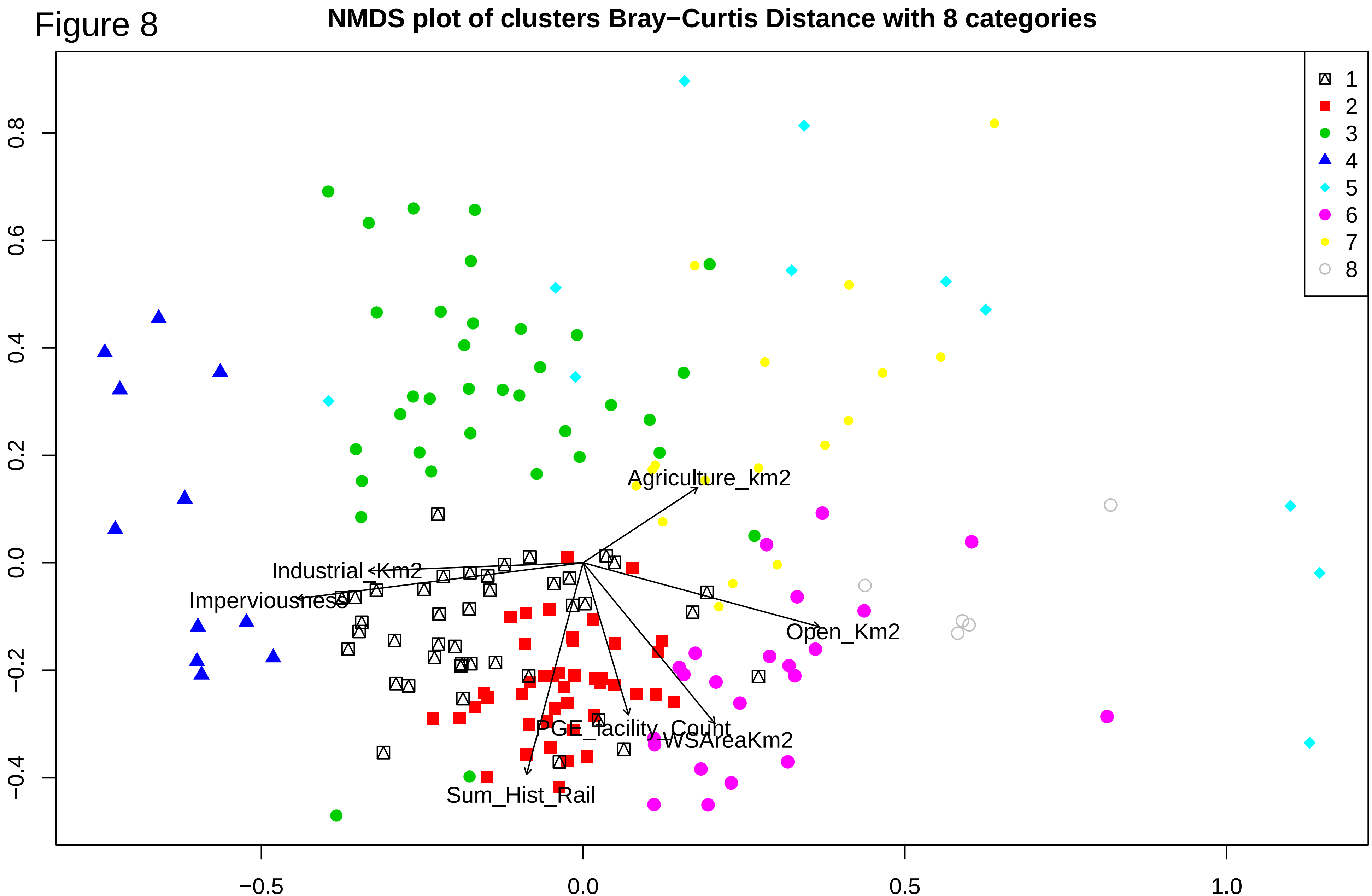


Figure 7
Watersheds
Bray-Curtis Dissimilarity C
1
2
3
4
5
6
7
8
0 5 10 20
Kilometers
N

Figure 8



NMDS plot of clusters Bray−Curtis Distance with 8 categories
1
2
3
4
5
6
7
8
Agriculture_km2
Industrial_Km2
Imperviousness
Open_Km2
PGE_facility_Count
WSAreaKm2
Sum_Hist_Rail
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.5
0.0
0.5
1.0

Figure 9

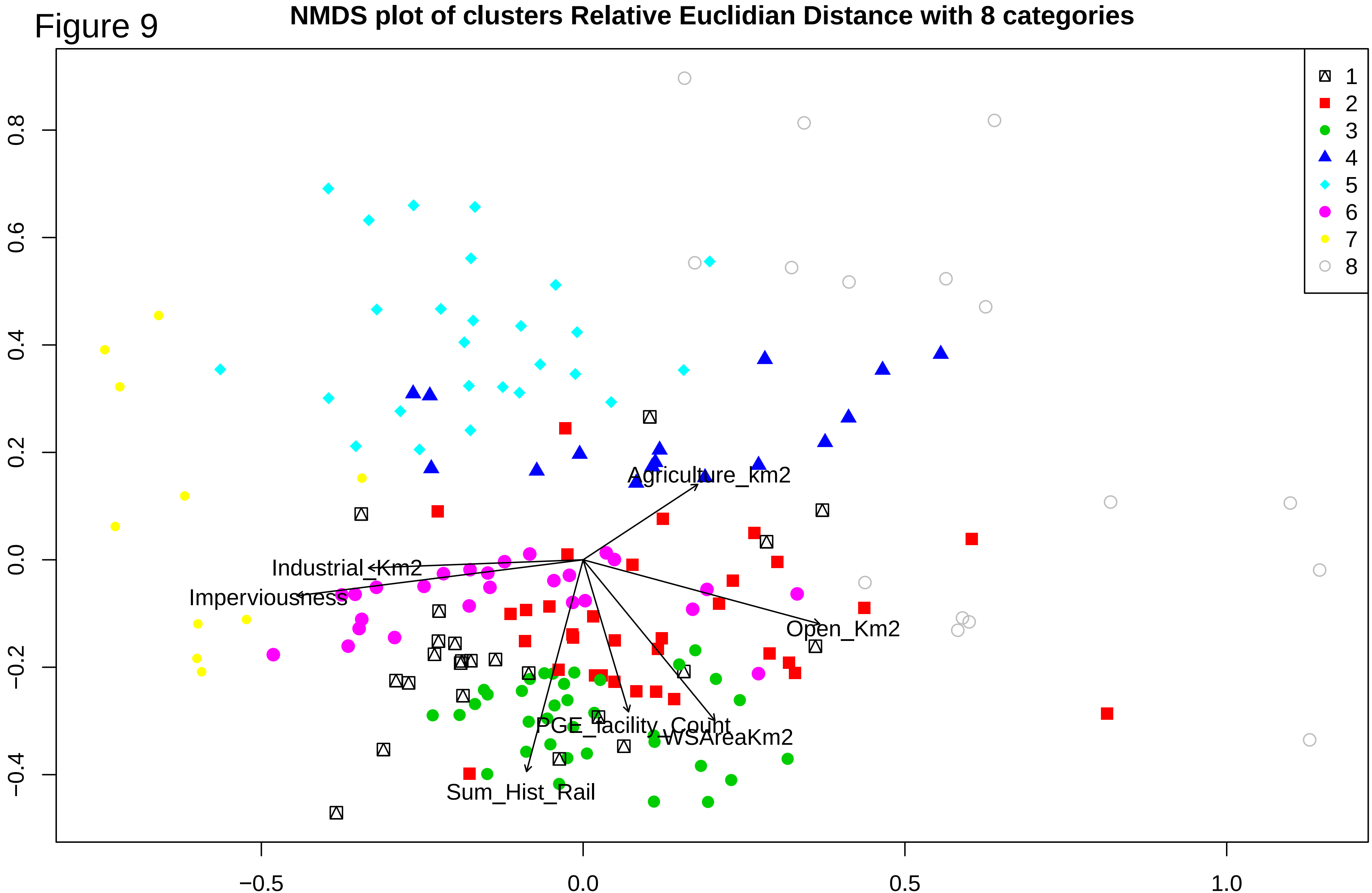

NMDS plot of clusters Relative Euclidian Distance with 8 categories
1
2
3
4
5
6
7
8
Agriculture_km2
Industrial_Km2
Imperviousness
Open_Km2
PGE_facility_Count
WSAreaKm2
Sum_Hist_Rail
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.5
0.0
0.5
1.0

Figure 10

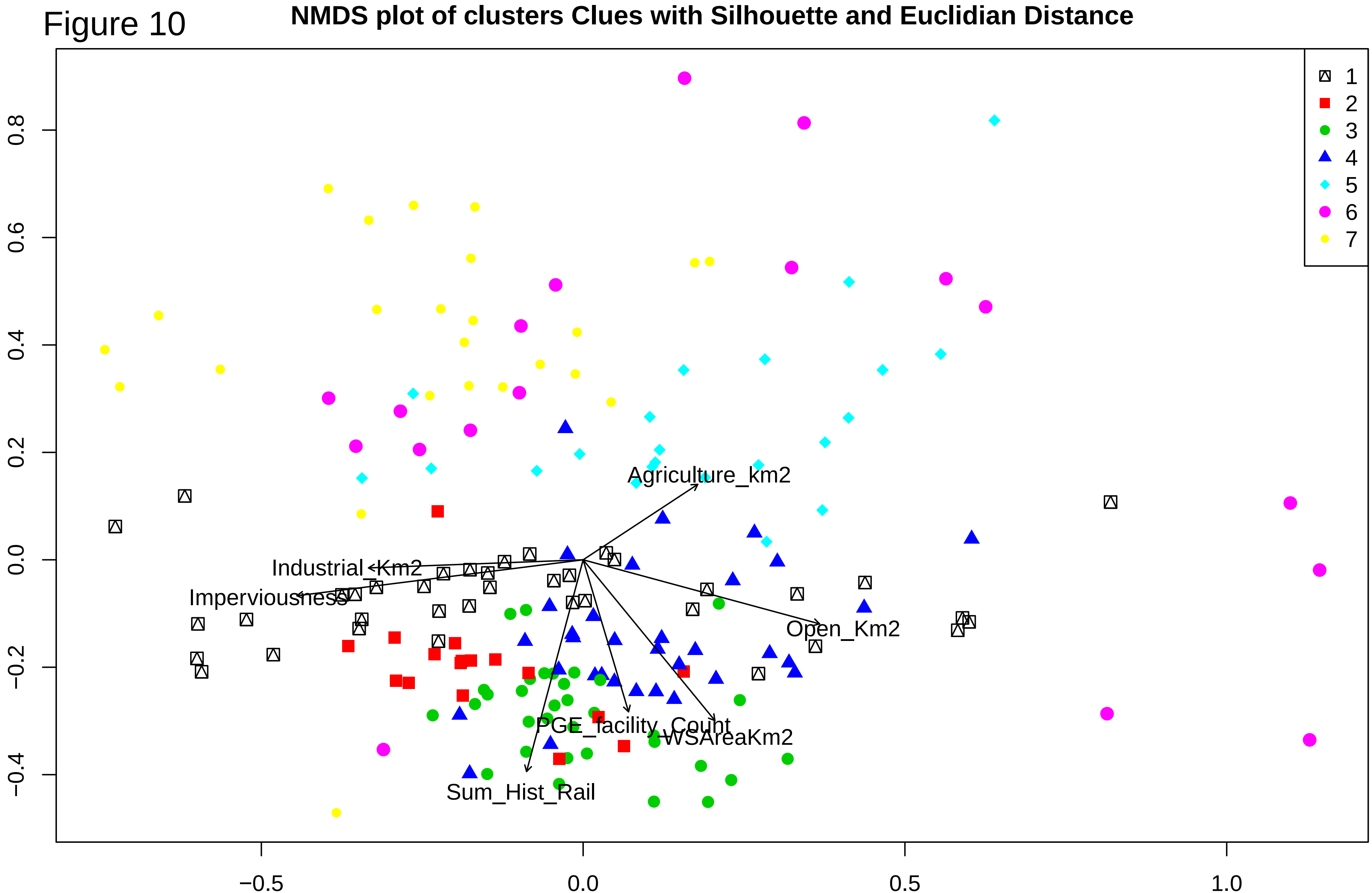

NMDS plot of clusters Clues with Silhouette and Euclidian Distance
1
2
3
4
5
6
7
Agriculture_km2
Industrial_Km2
Imperviousness
Open_Km2
PGE_facility_Count
WSAreaKm2
Sum_Hist_Rail
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.5
0.0
0.5
1.0

Figure 11

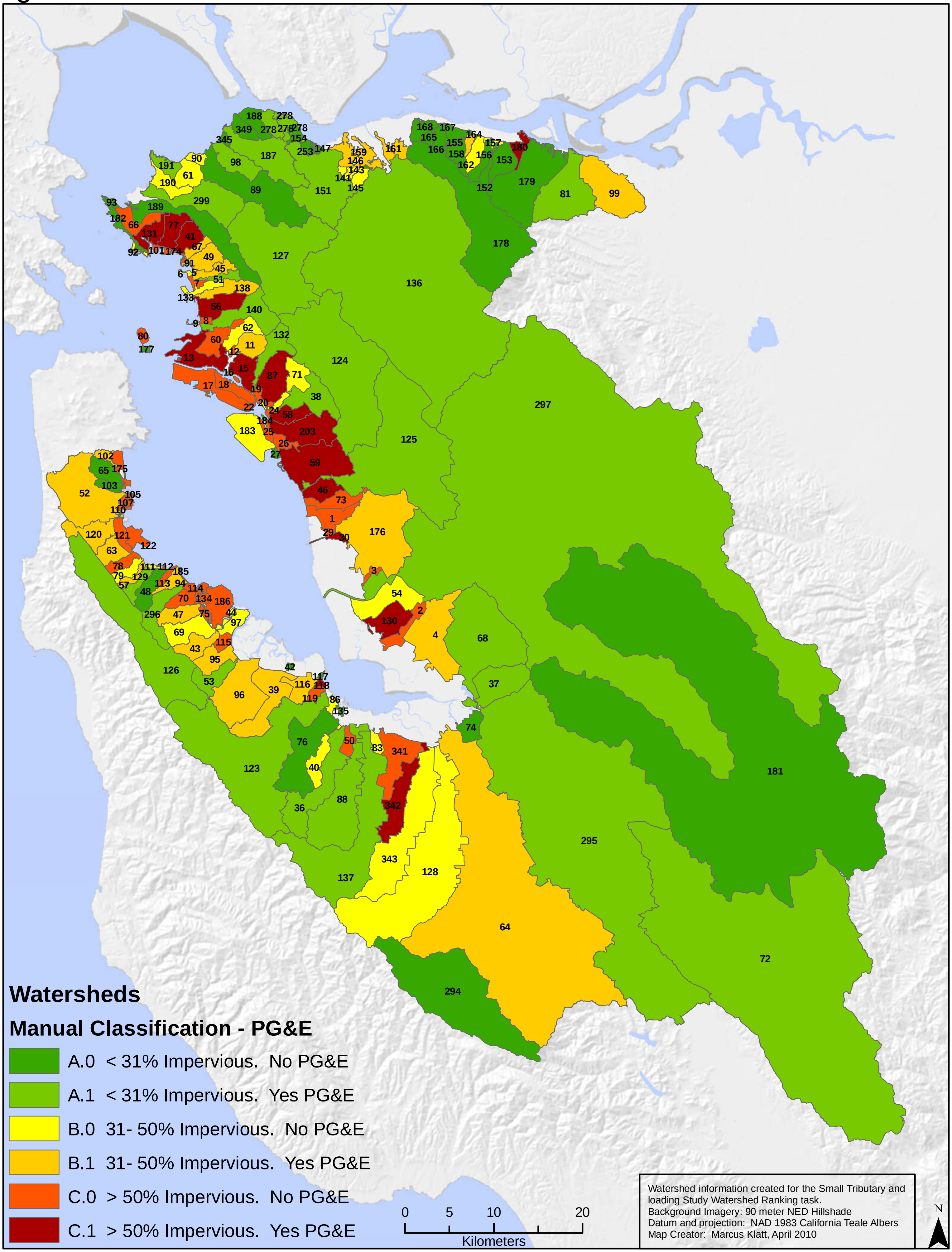
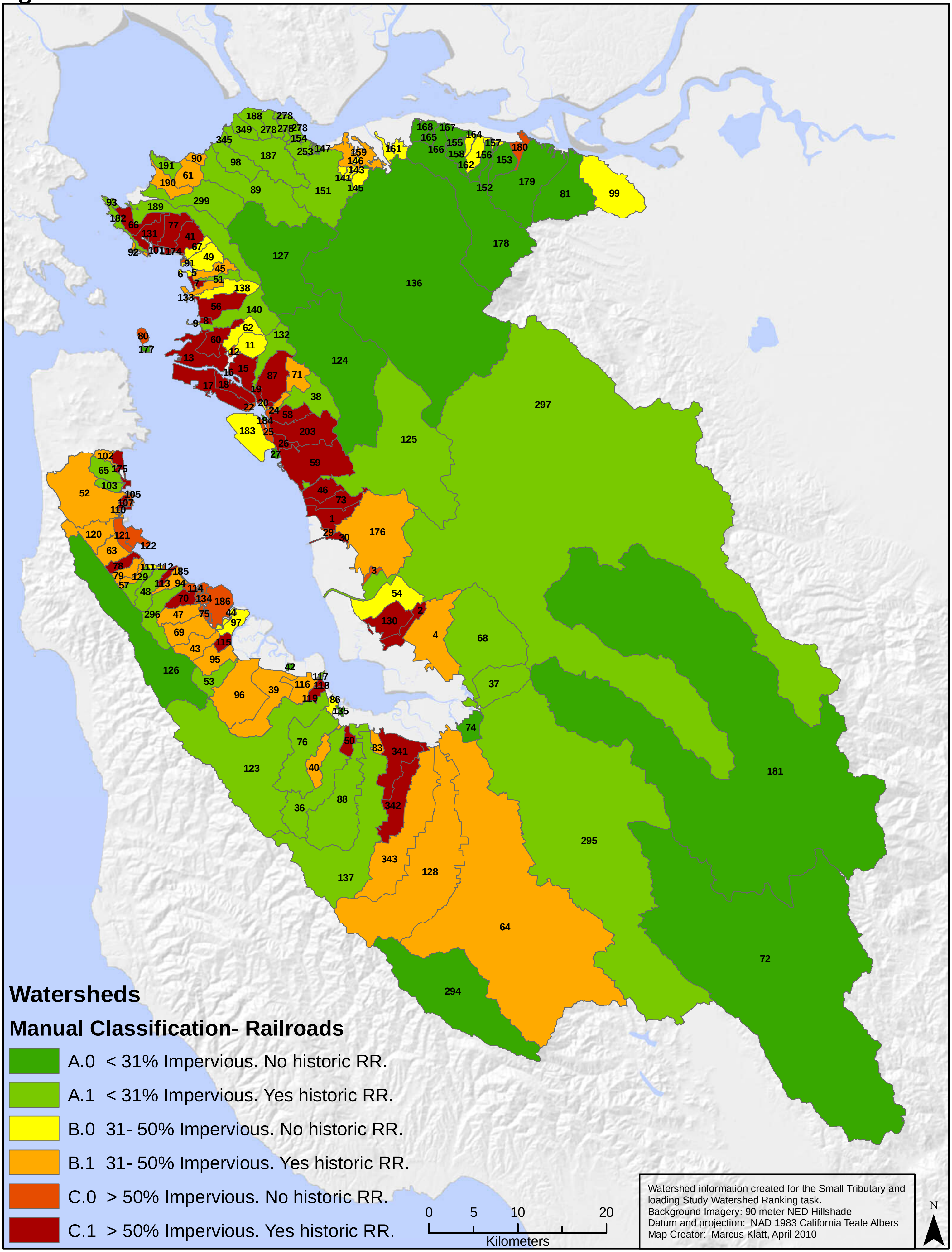

Watersheds
Manual Classification - PG&E
A.0 < 31% Impervious. No PG&E
A.1 < 31% Impervious. Yes PG&E
B.0 31- 50% Impervious. No PG&E
B.1 31- 50% Impervious. Yes PG&E
C.0 > 50% Impervious. No PG&E
C.1 > 50% Impervious. Yes PG&E
0 5 10 20
Kilometers
Watershed information created for the Small Tributary and loading Study Watershed Ranking task.
Background Imagery: 90 meter NED Hillshade
Datum and projection: NAD 1983 California Teale Albers
Map Creator: Marcus Klatt, April 2010
N

Figure 12


Watersheds
Manual Classification- Railroads
A.0 < 31% Impervious. No historic RR.
A.1 < 31% Impervious. Yes historic RR.
B.0 31- 50% Impervious. No historic RR.
B.1 31- 50% Impervious. Yes historic RR.
C.0 > 50% Impervious. No historic RR.
C.1 > 50% Impervious. Yes historic RR.
0 5 10 20
Kilometers
Watershed information created for the Small Tributary and loading Study Watershed Ranking task.
Background Imagery: 90 meter NED Hillshade
Datum and projection: NAD 1983 California Teale Albers
Map Creator: Marcus Klatt, April 2010
N