SANITATION DISTRICTS OF LOS ANGELES COUNTY

Converting Waste Into Resources

Robert C. Ferrante
Chief Engineer and General Manager
1955 Workman Mill Road, Whittier, CA 90601-1400
Mailing Address: P.O. Box 4998, Whittier, CA 90607-4998
(562) 699-7411 • www.lacsd.org

March 30, 2023

VIA ELECTRONIC MAIL

Ms. Renee Purdy, Executive Officer
California Regional Water Quality Control Board
Los Angeles Region
320 W. 4th St., Suite 200
Los Angeles, CA 90013

Joint Water Pollution Control Plant
CI No. 1758; Resolution R21-005; NPDES No. CA0053813
The Use of Chemical Scans to Characterize Toxic Effluent Samples (JWSS-21-005)
Special Study Final Report, March 2023

As required under Resolution R21-005, please find enclosed the final report for the subject special study. I certify under penalty of law that this document and all attachments were prepared under my direction or supervision in accordance with a system designed to assure that qualified personnel properly gather and evaluate the information submitted. Based on my inquiry of the person or persons who manage the system, or those persons directly responsible for gathering the information, the information submitted is, to the best of my knowledge and belief, true, accurate, and complete. I am aware that there are significant penalties for submitting false information, including the possibility of fine and imprisonment for knowing violations.

Very truly yours,

Joshua Westfall
Senior Environmental Scientist
Reuse and Compliance Section

JW
Enclosure

DOC# 6877795

Joint Water Pollution Control Plant

CI No. 1758; Resolution R21-005; NPDES No. CA0053813

The Use of Chemical Scans to Establish Chemical Baselines in Effluent (JWSS-21-005)

Final Report – March 2023

LOS ANGELES COUNTY
SANITATION DISTRICTS
Converting Waste Into Resources

Executive Summary

This study was initiated to build upon a previous study which chemically characterized non-toxic effluents from four Los Angeles County Sanitation Districts water reclamation plants (WRP). For this study, non-toxic samples for an additional WRP were characterized, and all toxic effluents from the previous study were characterized. Inspired by some groundbreaking non-targeted analyses, the goals of this study were to 1) assess dissimilarity among samples using cluster analysis 2) identify defining patterns such as variance by facility or season, and 3) identify defining patterns associated with toxic samples. Although toxicity at these facilities is relatively uncommon, investigating the occurrence of toxicity often takes considerable resources and doesn't always identify a causative agent. Despite the successful sampling, toxicity testing, and extensive chemical analyses, only the first goal was completely accomplished with some progress on the second. Over 160 final effluent samples were collected, scanned, and separated into clusters based on similarity in pesticide and metals scan results. Previous results suggested that low-resolution, semi-quantitative, chemical scans can provide sufficiently robust results to create a suite of chemical characteristics of non-toxic effluent and potentially, to differentiate between WRPs. Ultimately, toxic samples could not be effectively identified through this approach.

Background and Introduction

Historically, toxicity has been relatively infrequent in Los Angeles County Sanitation Districts (Sanitation Districts) final effluent samples, with only about 15% of final effluent toxicity tests being identified as "toxic" since the Test of Significant Toxicity (TST) was implemented in permits. Efforts to identify the cause(s) of toxicity are typically resource-intensive and have rarely been successful due to the transient and episodic characteristic of occurrence. Toxic samples have historically been investigated using a toxicity identification evaluation (TIE) approach to characterize toxicant(s) contributing to observed effects. If toxicity is transient or at low levels, this approach is especially challenging due to the lack of repeatability often encountered and the rapid disappearance of toxicity, typically within days of occurrence. However, chemical analyses can also be used to supplement or confirm results from TIE studies by ruling out common chemical compounds that are difficult to isolate through TIE manipulations, and properly stored samples can be analyzed well after the toxicity test has been completed.

Chemical testing for toxicity identification has historically focused on targeted analyses (e.g., the presence and concentrations of specific, commonly-encountered compounds such as ammonia, metals, and pesticides). This approach relies upon previously identified toxic thresholds for individual analytes or groups with similar modes of action (i.e., those expected to have additive effects). Thresholds are typically determined through studies looking at only a single toxicant in a far less complex matrix than WRP effluent. In addition, the selection of the targeted analytes is limited by knowledge bias and tends to focus on previously identified "problem" chemicals.

A more recent approach removes the knowledge bias by using non-targeted analyses (NTA) to identify toxicants (e.g., Peter et al., 2018). In this approach, samples are scanned for a broad range of signals corresponding to both known and unknown compounds, thereby creating a chemical "fingerprint." A clustering analysis is then applied to the resulting data, to quantify the magnitude of the similarities between profiles (i.e., results from all chemical analyses performed) and identify groups or clusters that share similar characteristics. The number of clusters in an analysis can vary between one (all samples in a single cluster) and the number of samples (each sample in its own cluster). Neither of these extreme cases is particularly informative to identify toxicants; the former provides no differentiation because all samples are considered

similar, while the latter provides no insights because each sample is unique. An "optimal" number of clusters balances compression (creating fewer clusters to group more results together) and accuracy (creating more clusters to differentiate among results), and groups samples whose characteristics are distinct from others, ideally based on some known feature such as season or location. However, this "optimal" number is somewhat arbitrary and varying based on the end use. Hierarchical clustering analysis (HCA) is a commonly used method, because the number of clusters in the analysis do not need to be specified beforehand and is, therefore, determined based on the specific patterns observed in the data. HCA defines similarity based on Euclidean distance, which is the mathematical "distance" between two data points. In this case, the distance between samples is calculated using the differences in the measured concentration values for all analytes (i.e., the differences between the complete chemical profiles of the samples). The shorter the Euclidean distance, the more similar two points or samples are to each other; groups of data points that are close together form clusters of samples that share similar chemical profiles. Ideally, these clusters would associate with biological impacts observed to help guide focused toxicity investigations.

NTA offers clear advantages over traditional toxicity identification methods, since it analyzes both known and unknown compounds, and the sources of toxicity are often unknown. However, it does require development of baseline chemical scans of non-toxic samples for comparison with toxic sample profiles. In addition, it typically uses highly specialized instrumentation that can detect species at very low concentrations while producing very high-resolution data. This type of instrumentation is relatively expensive and is not readily available in many wastewater laboratories.

This study employed the hybrid approach identified in a 2019 study, through the use of broad spectrum chemical scans. The goal was to continue development of a baseline for non-toxic samples and determine if this approach could identify toxic samples. Analyzing this broad range of constituents allows for examination of potential toxicants beyond those typically tested during toxicity investigations, and the associated clustering analysis could enable identification of similarities in effluent by plant or by season. If successful, this approach might represent a more economical starting point for less knowledge-biased chemical investigations.

METHODS

Sample Collection

The samples analyzed for this study were collected for toxicity testing under National Pollutant Discharge Elimination System (NPDES) permits. Since the goal of this study was to associate chemical fingerprints to differentiate between toxic and non-toxic samples, potentially identifying specific analytes associated with toxicity (as defined using standard USEPA whole effluent toxicity testing protocols and the TST statistic), sample collection deviated significantly from what would normally be used for chemical analyses. Flow-weighted 24-hour composite samples were used in all cases. These composite samples were collected in non-air-tight containers and then poured into low-density polyethylene (LDPE) containers. A typical toxicity test uses three samples, with each being used for one to three days. Previous efforts (focused on sample collection and analytical results), as well as the methods employed in the toxicity tests, strongly suggested that volatile compounds were unlikely to be present in measurable concentrations; these compounds were therefore excluded from analysis for this study. Chemical screening included a metals scan and a current-use pesticide suite. A semi-volatile screen was evaluated in the previous study but not used in this iteration.

The five WRPs shown in Table 1 were sampled at least three times monthly for one year, from July 2019 to June 2020 for Los Coyotes, Whittier Narrows, San Jose Creek, and Saugus and from July 2021 to June 2022 for Pomona. These WRPs were selected to represent the diversity in size, influent source, and treatment methods and some history of observing toxicity.

Table 1. Study WRPs

WRP	NPDES No.	Characteristics
Los Coyotes (LC)	CA0054011	Step-feed activated sludge treatment facility with chlorine-based disinfection, lowest frequency of observed toxicity and highest industrial influent base for WRPs at 14%
Whittier Narrows (WN)	CA0053716	Small Modified Ludzack-Ettinger (MLE) activated sludge treatment facility with UV disinfection
Pomona (PO)	CA0053619	Small Modified Ludzack-Ettinger (MLE) activated sludge treatment facility with chlorine-based disinfection
San Jose Creek East (SJ)	CA0053911	Step-feed activated sludge treatment facility with chlorine-based disinfection, higher frequency of observed toxicity
Saugus (SA)	CA0054313	Small Modified Ludzack-Ettinger (MLE) activated sludge treatment facility with chlorine-based disinfection

Whole Effluent Toxicity Testing

Each WRP monitored in this study used the water flea, *Ceriodaphnia dubia,* as the chronic toxicity test species. Tests were conducted according to the *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* (USEPA 2002). *Ceriodaphnia dubia* were exposed in a static renewal system to different concentrations of effluent for a period of six to eight days. Tests were terminated on the day when at least 60% of the surviving female control organisms had three broods or eight days after test initiation, whichever occurred first. Test results were based on survival and reproduction; sub-lethal (i.e., reproduction) toxicity was determined using the Test of Significant Toxicity (TST) test statistic, as required by each WRP's NPDES permit. All method-defined Test Acceptability Criteria (TAC) were met or the toxicity test was considered invalid and repeated. TAC include the requirements that there must be at least 80% control survival and the surviving control organisms produce at least an average of 15 or more young in the three broods.

Chemical Scans

Broad-spectrum scans were conducted for 85 analytes using three analytical methods. A custom high-performance liquid chromatography (HPLC) scan was utilized for current use pesticides (Appendix A). Samples were scanned for metals using inductively coupled plasma mass spectrometry (ICP-MS) and the semi-quantitative feature of the instrument software to give estimates of an extended analyte list (Appendix A) that is not limited to the analyte lists found in the Sanitation Districts' fully quantitative metals methods (e.g., EPA 200.8). All analyses were conducted by the Sanitation Districts' San Jose Creek Water Quality Laboratory. Sampling and analysis began in July 2019 and ended in June 2022.

Data Analysis

All raw data are maintained within both the Sanitation Districts' document management system database and Laboratory Information Management System. All data analysis was conducted using the R statistical computing platform. HCA was performed to compute the Euclidean distance between samples, using analyte concentrations and the HCUT function in the factoextra R package.

Clusters were developed using complete and centroid methods, then evaluated for method differences. HCA does not require a specific number of clusters; instead, a statistically "optimal number of clusters" was previously identified using the "fviz_nbclustering" function. Use of this function provided an objective and statistically meaningful approach to select cut points (i.e., using established algorithms to balance between maximum compression and accuracy) along the resulting dendrograms, which was supplemented with professional judgement. This analysis was conducted using the gap statistic (Tibshirani et al. 2001), elbow method (based upon the total within-cluster sum of squares as a function of the number of clusters), and silhouette (Kaufman and Rousseeuw, 1990) approaches for comparison. Although 2 and 10 clusters were identified using differing techniques, a post-hoc HCA was also conducted using five clusters due to its relevance as an intermediate between the identified cluster sizes and its relevance to suspected contributing factors (i.e., seasonality and WRP).

Once a target number of clusters was identified, clustering analyses were finalized and the resulting dendrograms were created using the dendextend package. Select analytes were excluded from the clustering analysis. Specifically, carbon was excluded because its high concentration in all samples might effectively quench other signals; nitrogen and chlorine were also censored due to high concentrations (relative to most analytes) and the fact that all samples were dechlorinated before exposing test organisms. Finally, select rare-earth metals were added to the analyte list in January 2020 but were excluded from data analysis to maintain a consistent analyte list.

RESULTS

Sample Collection and Analysis

During this study, there were no sampler failures at any of the five, targeted WRPs during the study term; the sampling success was 100%, exceeding the Sanitation Districts' Receiving Water Quality Assurance Project Plan goal of 90%. In total, 168 samples were successfully analyzed for chemical fingerprints, varying among facilities, the number of samples varies among the plants. Analytes that had no detections in any samples (amphetamine, spiroxamine, malathion, dimethoate, dichlorvos, spirotetramat, holmium, tantalum, and rhodium) were also excluded from the data analysis in an effort to reduce cluster skew.

Whole Effluent Toxicity Testing

Because each toxicity test requires at least three samples, there are far more samples than toxicity test results. Overall, the observed frequency of toxicity during this study was comparable to historical results from 2007 to 2022, apart from the San Jose Creek East WRP, where a majority of toxicity tests during the 2019 – 2020 sampling exhibited toxicity.

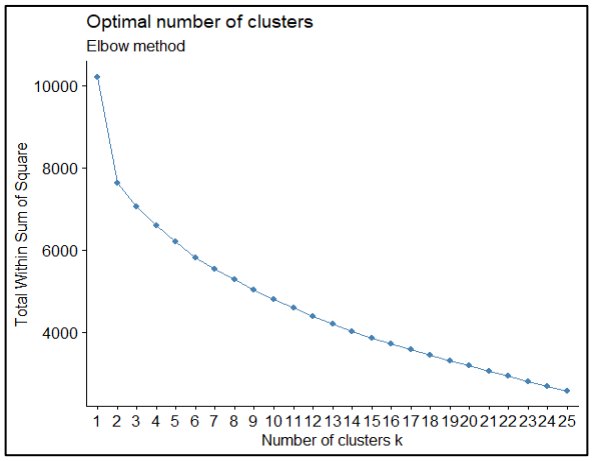
Discussion

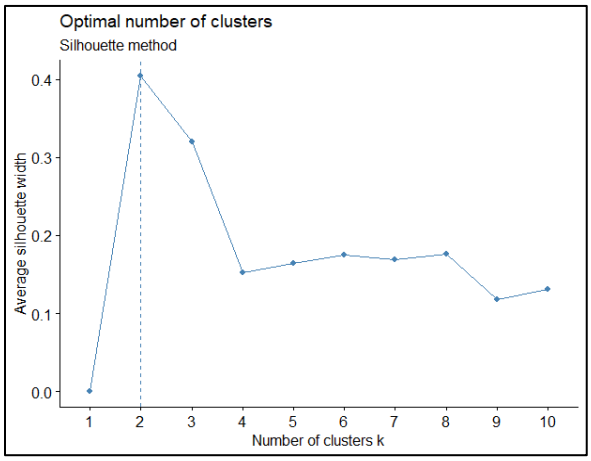
Based upon the three cluster optimization models (gap statistic, elbow, and silhouette), two optimal cluster sizes were identified. The elbow method was unable to identify an optimal number of clusters and was not utilized further; the silhouette method identified two as the optimal number of clusters, and the gap statistic identified ten clusters as the optimum (Figures 1a, 1b, 1c).

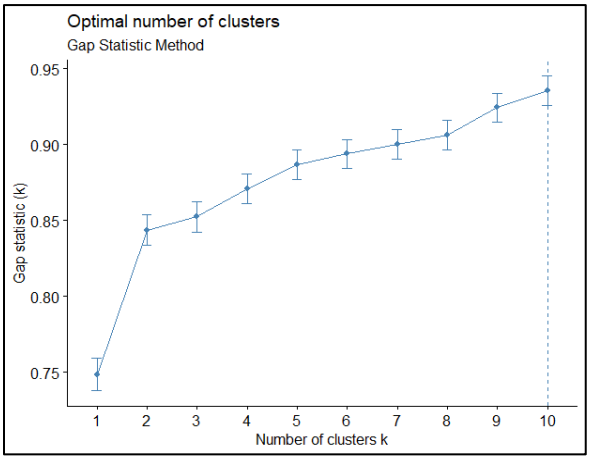
Euclidean distance based, hierarchical clustering was conducted on the analytical chemistry results using two, five, and ten clusters. Figure 2 shows a color-coded dendrogram for the most conclusive (ten-cluster) analysis, and the accompanying summary tables (Table 2) shows the cluster breakdown by WRP. No association was observed between month/season and clustering. The two- and five-cluster models were also analyzed but did not yield insights beyond the ten-cluster model (Appendix B).

In the dendrogram analysis, distances between clusters represent similarity in chemical characteristics; clusters whose junction on the vertical axis are farther apart are less similar. To identify unique chemical characteristics among clusters, outliers (i.e., clusters on the edges of the dendrogram) were analyzed in the two-, five-, and ten-cluster models; however, no patterns by WRP, toxicity, or season were found.

Figure 1. Cluster Optimization Output. a: Elbow Method; b: Silhouette Method; c: Gap Statistic Method (Dashed vertical line represents optimal number of clusters as determined by the algorithm)

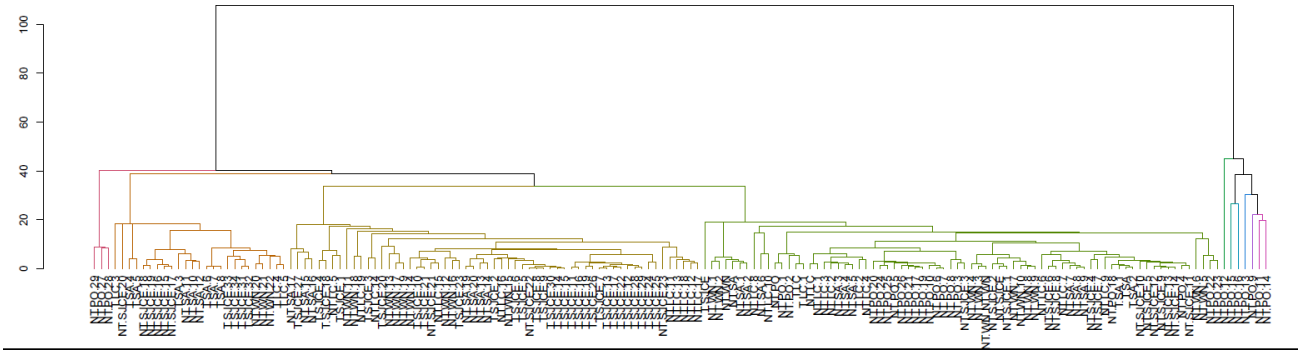






As shown in Figure 2 and Table 2, there was little recognized differentiation between WRPs and "noise" or outliers. Pomona effluent were the only samples with recognizable spread across the clusters; most facilities were concentrated in one or two, with most of all samples being in cluster one. There was no systematic differentiation of "toxic clusters." This is likely based on a multitude of factors including the notion that toxicity is a strictly relative measure and often low-level, there is inherent biological variability and these analytes are often being detected near the reporting level where confidence is lower. Based on these samples, it does appear that effluent from different WRPs can be identified on many occasions with limited confidence using the analytical chemistry capabilities currently employed by the Sanitation Districts' laboratories. However, this approach will not work in its current format for identification of likely causal agents to any observed toxicity.

Figure 2. Hierarchical Clustering of Effluent Chemical Analyses Using Ten Clusters



To date, the specific chemical characteristics that define the clusters have not been identified. The relatively close distance between the WRP clusters on the dendrogram likely indicates that the chemical characteristics are relatively similar between these clusters, compared to outlying clusters. This similarity makes identification of the unique chemical signal of each cluster (possibly resulting from slight variations among multiple analytes) more difficult. Further analyses of the data may help differentiate the characteristics among WRP clusters, but may be limited by the relatively small analyte list compared to high-resolution NTA scans reported in the literature (e.g., Ulrich et al. 2019), the number of samples analyzed, and low levels of some compounds in the broad spectrum scans (i.e., number of non-detected analytes).

Table 2. Ten-Cluster Population by WRP

Cluster Population by WRP (10 Clusters)					
Cluster	**LC**	**PO**	**SA**	**SJCE**	**WN**
1	13	20	13	16	12
2	0	1	0	0	0
3	0	2	0	0	0
4	0	1	0	0	0
5	0	1	0	0	0
6	0	1	0	0	0
7	0	1	0	0	0
8	0	3	0	0	0
9	10	0	7	33	9
10	3	0	9	10	3

CONCLUSIONS

The first goal of this study was achieved: broad spectrum, semi-targeted chemical scans utilizing analytical methods currently employed by the Sanitation Districts' laboratories were applied to non-toxic and toxic effluent samples. The second goal, use of these scans to focus efforts while investigating toxicity, was not successful as there were no clusters of predominantly toxic samples and success was limited in the use of analysis to identify clusters of samples with similar chemical characteristics based on WRP. No such fingerprints were identified based on seasonality (i.e., by month) using the two, five, or ten cluster models. In addition, most samples were in a single cluster, reducing the value.

In analyzing the data, the tabulation and visualization of analytical output were found to be both complementary and necessary to identify relevant patterns. Dendrograms allowed visual identification of clusters but tabulation clearly delineated the strong clustering by WRP, thus demonstrating the importance of presenting results in multiple formats.

This study also revealed some potential limitations of this approach: (1) the specific chemical characteristics that define the WRP clusters have not yet been identified, and (2) the ability to resolve the differences between samples may be limited by the relatively small analyte list and sample size, and the low levels of some compounds in the broad spectrum scans (i.e., number of non-detected analytes). However, continued advances in methodology and instrumentation technology are expected to improve the resolution on this type of analysis, particularly for issues associated with non-detections.

In summary, this study provides evidence that toxicity in samples from WRPs cannot be differentiated based on chemical characteristics. Similar to a traditional TIE approach, it appears that the higher the magnitude and frequency of an effect, the easier it would be to characterize. Due to the lack of success, this effort is unlikely to be pursued further at this time in a predictive format. However, these scans will continue have a role in TIE efforts to help rule toxicants out and find anomalously high concentrations of a range of analytes.

References:

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. " *Finding groups in data: an introduction to cluster analysis* 344: 68-125

Peter, Katherine T., et al. 2018 "Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon." *Environmental Science & Technology* 52.18 : 10317-10327.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2: 411-423.

Ulrich, E., J. Sobus, Chris Grulke, A. Richard, S. Newton, M. Strynar, K. Mansouri, AND A. Williams. 2019. "EPA's Non-Targeted Analysis Collaborative Trial (ENTACT): Genesis, Design, and Initial Findings." *Analytical and Bioanalytical Chemistry*.411(4):853-866.

U.S. Environmental Protection Agency. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* – Fourth Edition. Washington, D.C., October 2002. EPA/821-R-02-013.

Appendix A: Target Analytes

Current Use Pesticide Screen	Metals Scan	Metals Scan	Metals Scan
DEET	Beryllium	Selenium	Neodymium
Imidacloprid	Boron	Bromine	Samarium
Diuron	Sodium	Rubidium	Europium
Thiamethoxam	Magnesium	Strontium	Gadolinium
Acephate	Aluminum	Yttrium	Dysprosium
Naled	Silicon	Zirconium	Erbium
Chlorpyrifos	Phosphorous	Niobium	Thulium
Carbaryl	Sulfur	Molybdenum	Ytterbium
Carbofuran	Potassium	Rutherfordium	Lutetium
Diazinon	Calcium	Palladium	Hafnium
Acetamiprid	Titanium	Silver	Tungsten
Thiacloprid	Vanadium	Cadmium	Rhenium
Propoxur	Chromium	Tin	Osmium
Imazalil	Manganese	Antimony	Iridium
Metaxyl	Iron	Tellurium	Platinum
Methamphetamine	Cobalt	Iodine	Gold
Paclobutrazol	Nickel	Cesium	Mercury
Boscalid	Copper	Barium	Thallium
Micobutanil	Zinc	Lanthanum	Lead
Azoxystrobin	Gallium	Cerium	Bismuth
Piperonyl Butoxide	Arsenic	Praseodymium	Thorium
			Uranium

Appendix B. Additional Output

Figure B.1. Complete Hierarchical Clustering of Effluent Chemical Analyses Using Two Clusters

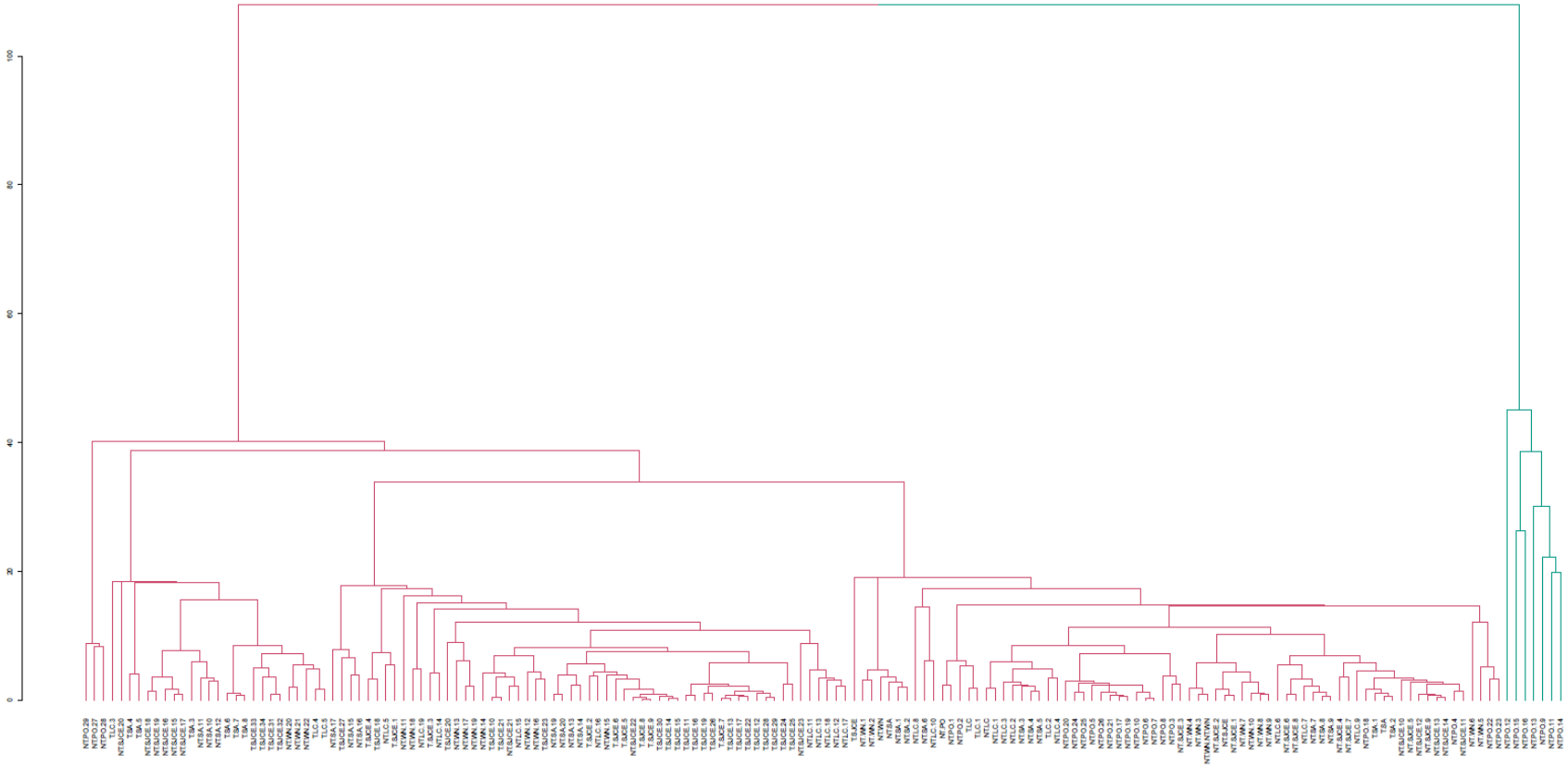


Table B.1. Two-Cluster Population by WRP

Cluster Population by WRP (2 Clusters)					
Cluster	**LC**	**PO**	**SA**	**SJCE**	**WN**
1	26	23	29	59	24
2	0	7	0	0	0

Figure B.2. Hierarchical Clustering of Effluent Chemical Analyses Using Five Clusters

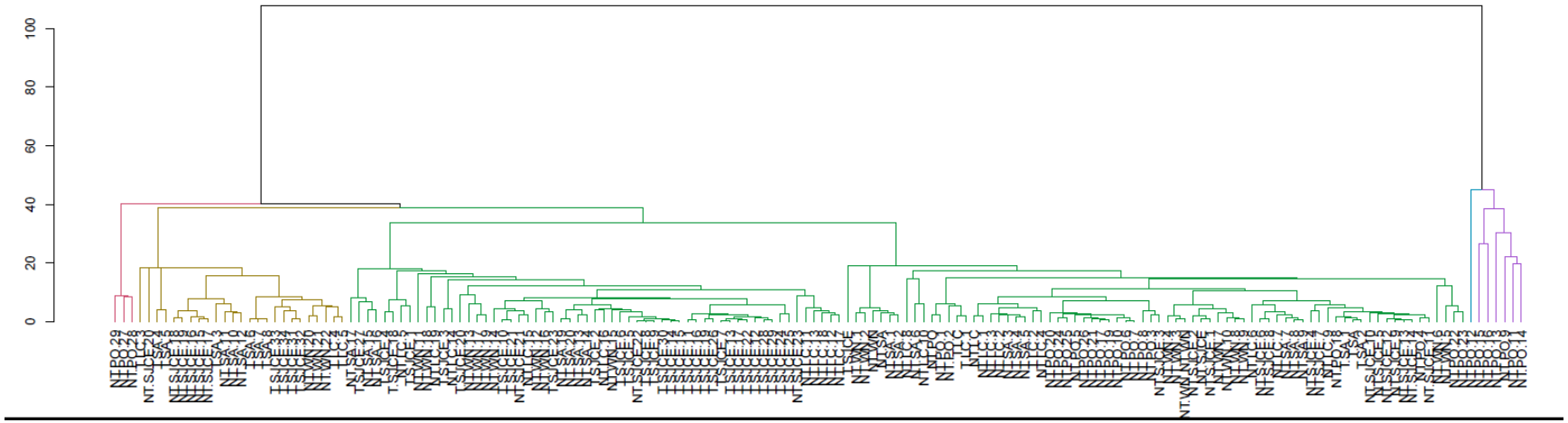


Table B.3. Five Cluster Population by WRP

Cluster Population by WRP (5 Clusters)					
Cluster	**LC**	**PO**	**SA**	**SJCE**	**WN**
1	23	20	20	49	21
2	0	6	0	0	0
3	0	1	0	0	0
4	0	3	0	0	0
5	3	0	9	10	3