

Application of Principal Component and Hierarchical Cluster Analysis to Delineate VOC Plumes at a Fractured Rock Site

**Tori Ward, Sercan Ceyhan, Dan Bryant, Jason House,
Sam Olney**

June 5, 2024



Two Suspected Source Areas

- Former Seepage Pit Area
 - Disposal for at least 15 years
- Chemical Laboratory

Primary chemicals of concern:

- TCE, 1,1,1-TCA, & degradation products
- 1,4-Dioxane



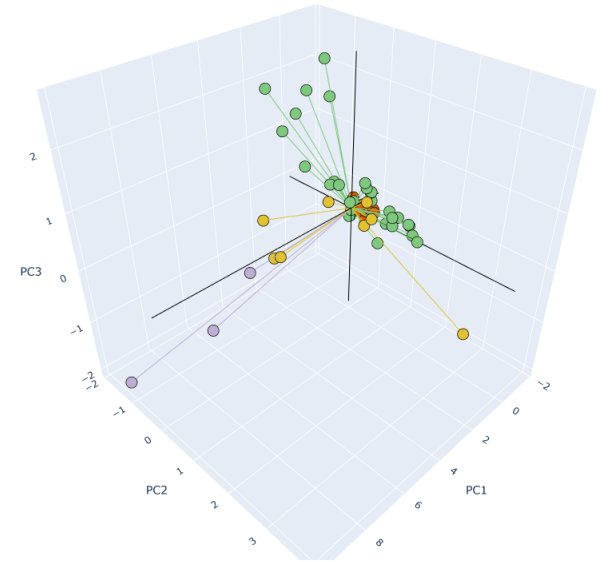
Objectives

- Desire to differentiate plumes
 - Can source and plume characteristics be statistically differentiated?
- Refine the CSM
 - Provide unbiased, quantitative lines of evidence
 - Determine if statistical results corroborate other analyses
- Apply principal component and hierarchical cluster analysis
 - Reduce a large data set to a smaller number of variables

Methodology - Intro to PCA and HCA

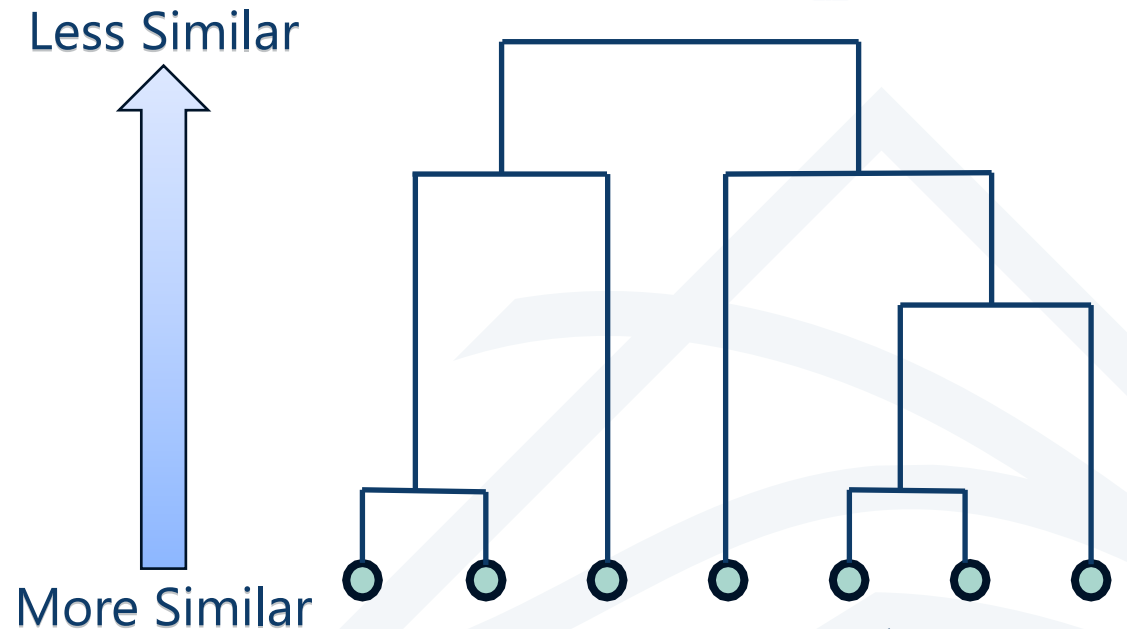
- **Principal Component Analysis (PCA)**

- Reduces many variables to a few components that explain most of the variability in the data
- **Components** are linear combinations of variables, ordered by decreasing importance



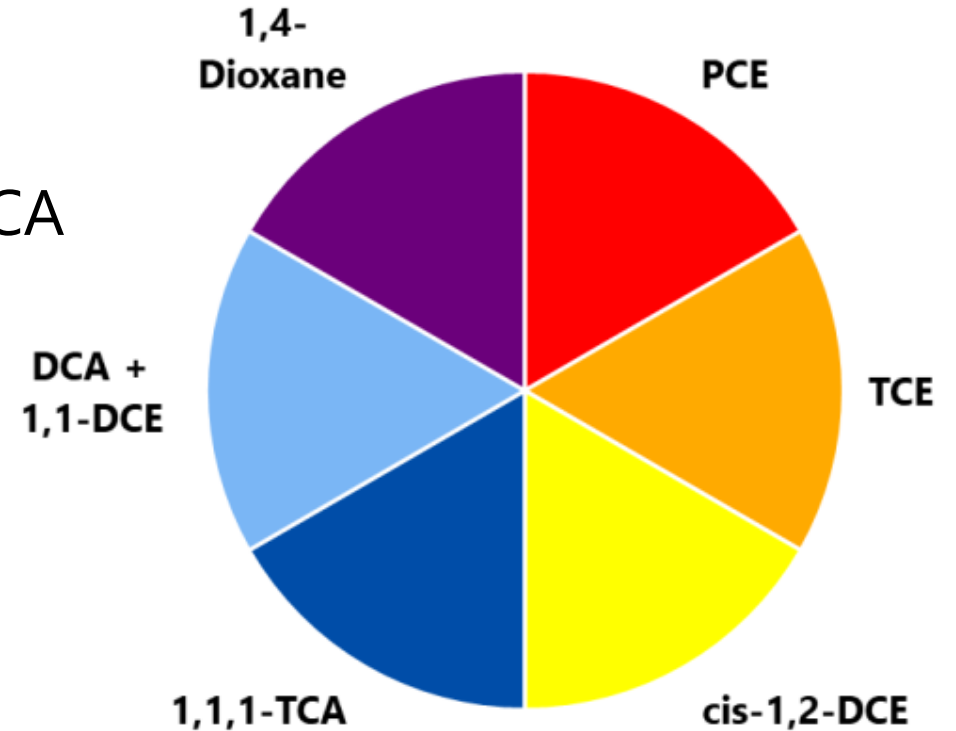
- **Hierarchical Cluster Analysis (HCA)**

- Starts by placing each observation in its own **cluster** (Bottom-up approach)
- Applies a distance measure to merge clusters that are most similar



Methodology – Input Data

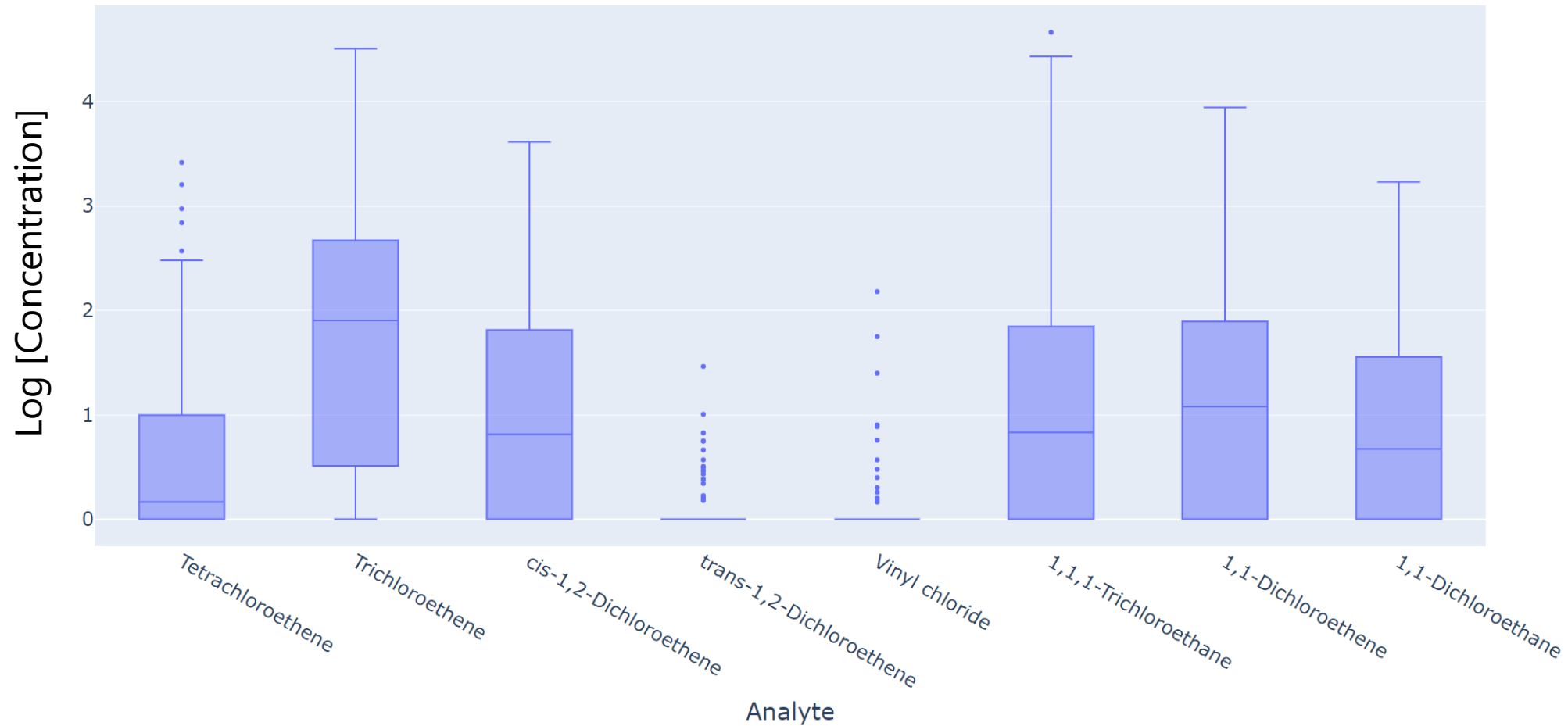
- Convert $\mu\text{g/L}$ to molarity
- Aggregate TCA degradation products 1,1-DCA & 1,1-DCE into one feature
- Exclude trans-1,2-DCE and VC (low concentrations)
- Log_{10} -transformation of data set
- Exclude wells in which VOCs were ND



Methodology – Analysis Steps

- Basic statistics
- Pearson Correlation matrix
- PCA
 - Review components and explained variance
- HCA
 - Elbow and Silhouette Charts
 - Dendrogram
- 2D & 3D visualization to assess corroboration between PCA & HCA results

Methodology – Basic Stats



Key Point: VC & trans-1,2-DCE excluded due to low concentrations.

Methodology – Pearson Correlation Matrix

	1,1,1-TCA	1,4-Dioxane	DCA + 1,1-DCE	PCE	TCE	cis-1,2-DCE
1,1,1-TCA	1					
1,4-Dioxane	0.19	1				
DCA + 1,1-DCE	0.88	0.23	1			
PCE	0.76	0.2	0.74	1		
TCE	0.56	0.17	0.57	0.25	1	
cis-1,2-DCE	0.52	0.26	0.75	0.48	0.28	1

Key Point: High correlation between certain analytes suggests they can be aggregated into a smaller number of components

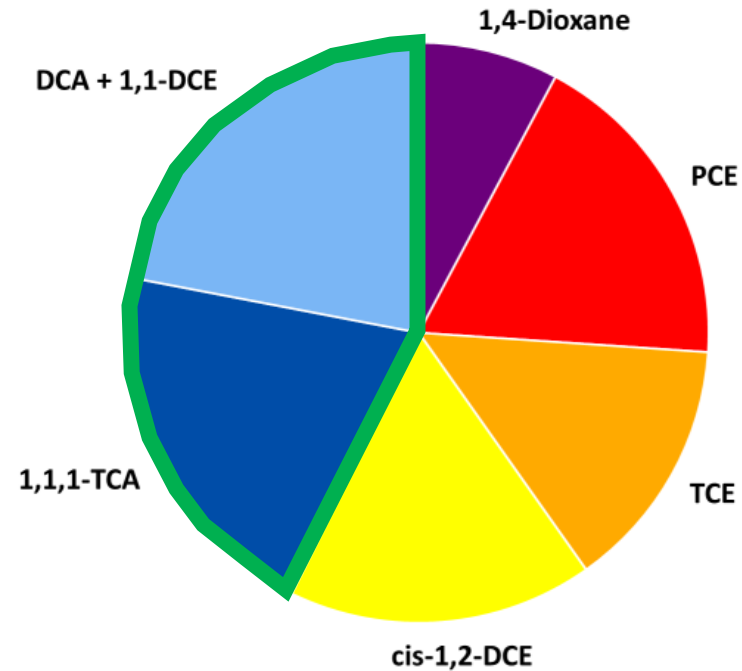
Methodology – PCA Component Weights

	1,4-Dioxane	PCE	TCE	cis-1,2-DCE	1,1,1-TCA	DCA + 1,1-DCE	Explained Variance
PC1	0.18	0.43	0.33	0.40	0.49	0.52	58.0%
PC2	0.95	-0.11	-0.13	0.14	-0.19	-0.10	15.6%
PC3	0.12	-0.41	0.85	-0.32	0.04	-0.04	13.3%
PC4	0.20	0.52	-0.09	-0.77	0.29	-0.12	9.4%
PC5	-0.05	0.61	0.38	0.15	-0.64	-0.24	2.9%
PC6	0.02	-0.05	-0.09	-0.33	-0.48	0.80	0.9%

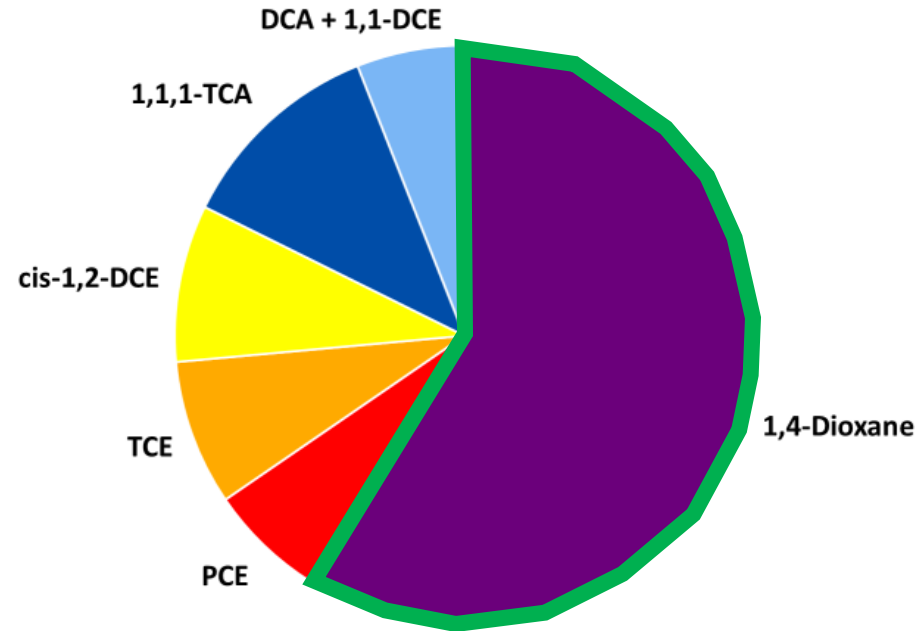
Key Point: 1,1,1-TCA and degradation products, 1,4-dioxane, and TCE explain most of the variability in the original data set.

Methodology – PCA Component Weights

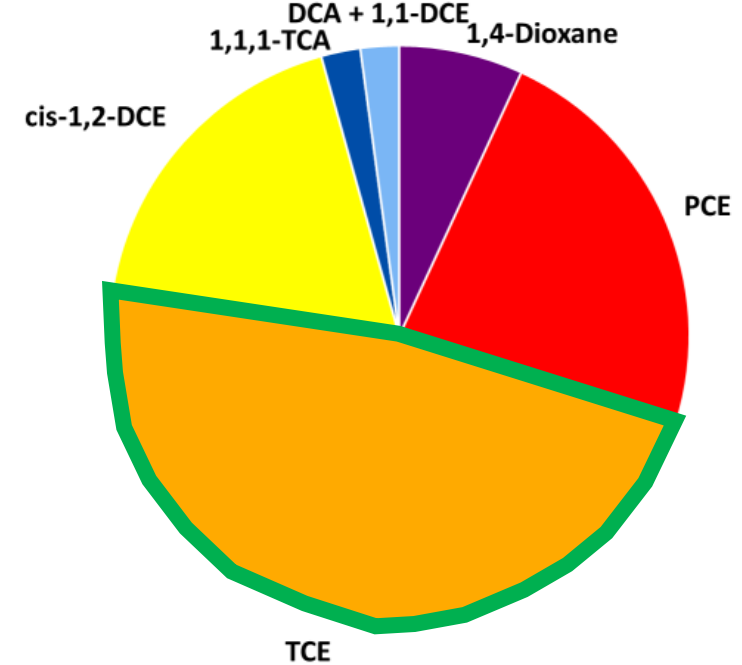
Principal Component 1



Principal Component 2



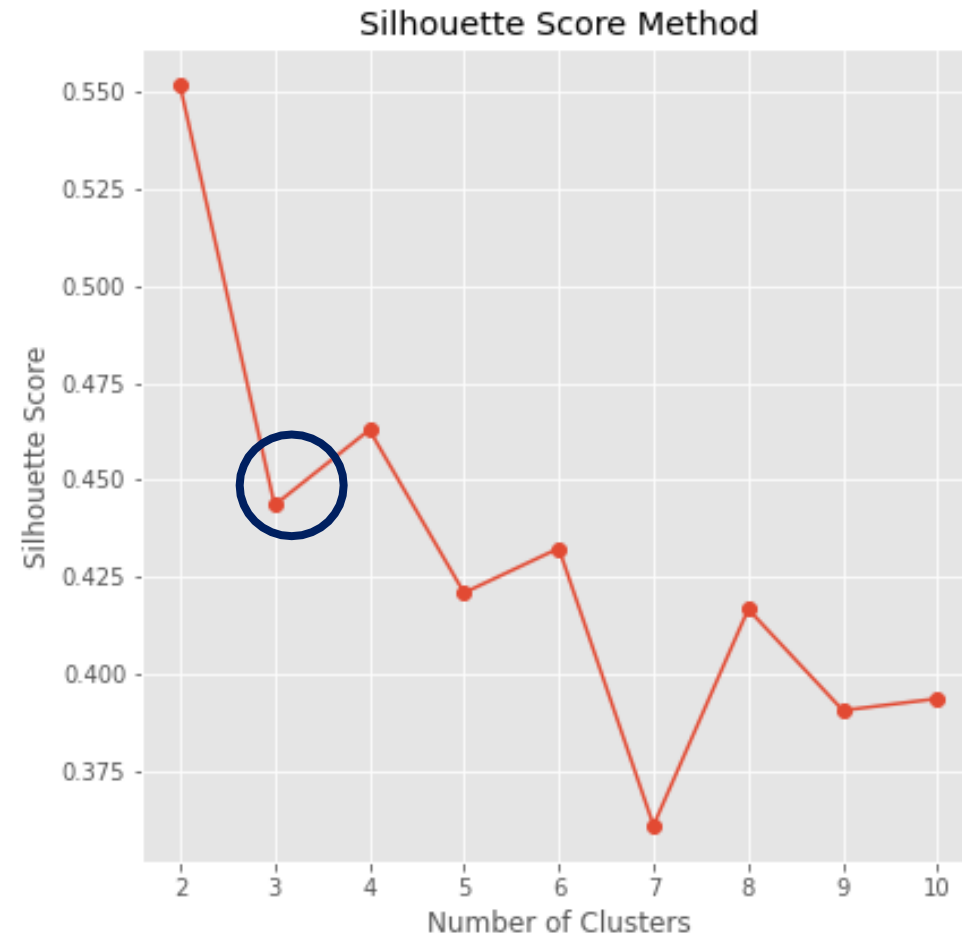
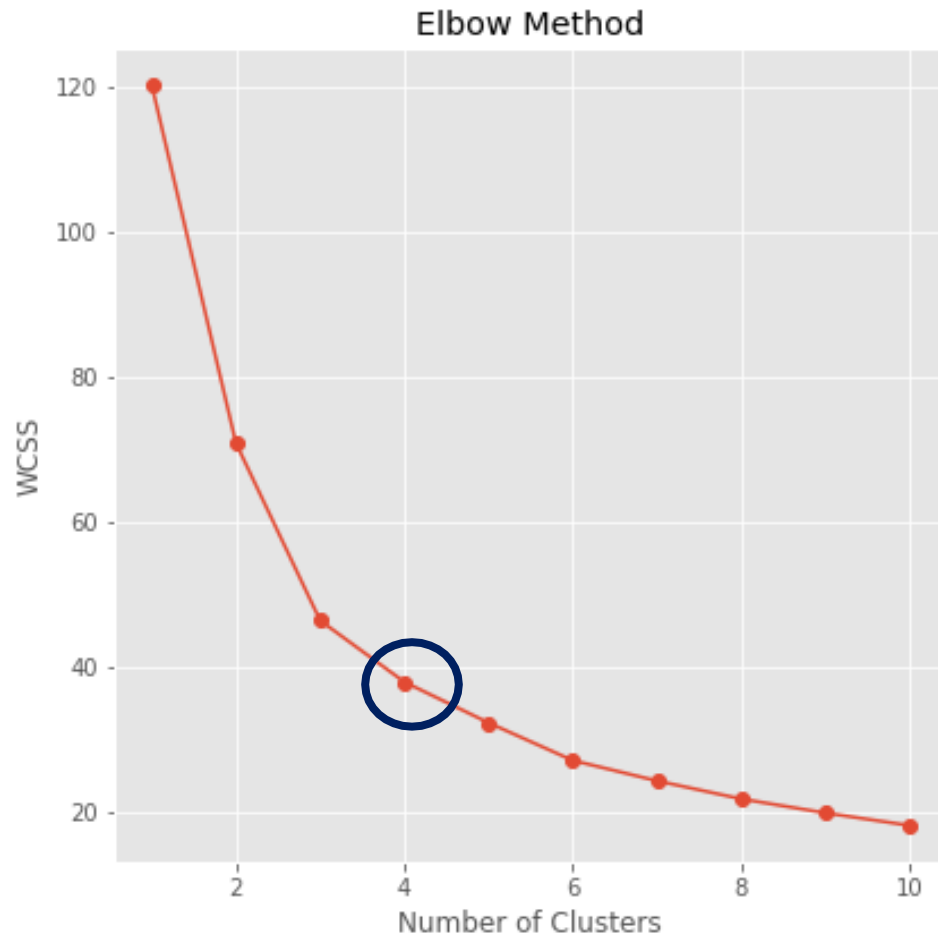
Principal Component 3



Key Points:

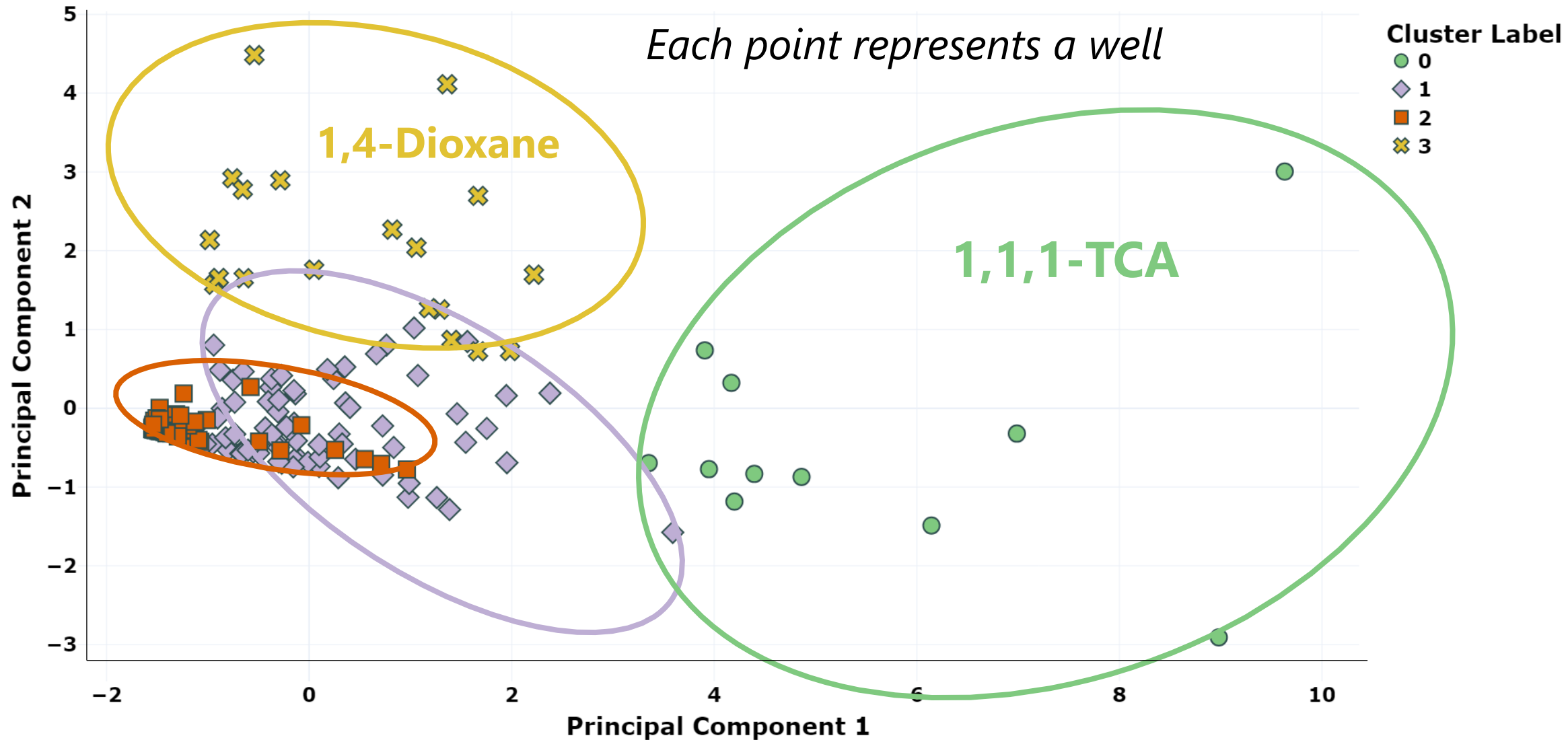
- Principal Component 1 is strongly tied to 1,1,1-TCA & degradation products
- Principal Component 2 is strongly related to 1,4-dioxane
- Principal Component 3 is strongly related to TCE

Results – Cluster Analysis Elbow and Silhouette Charts

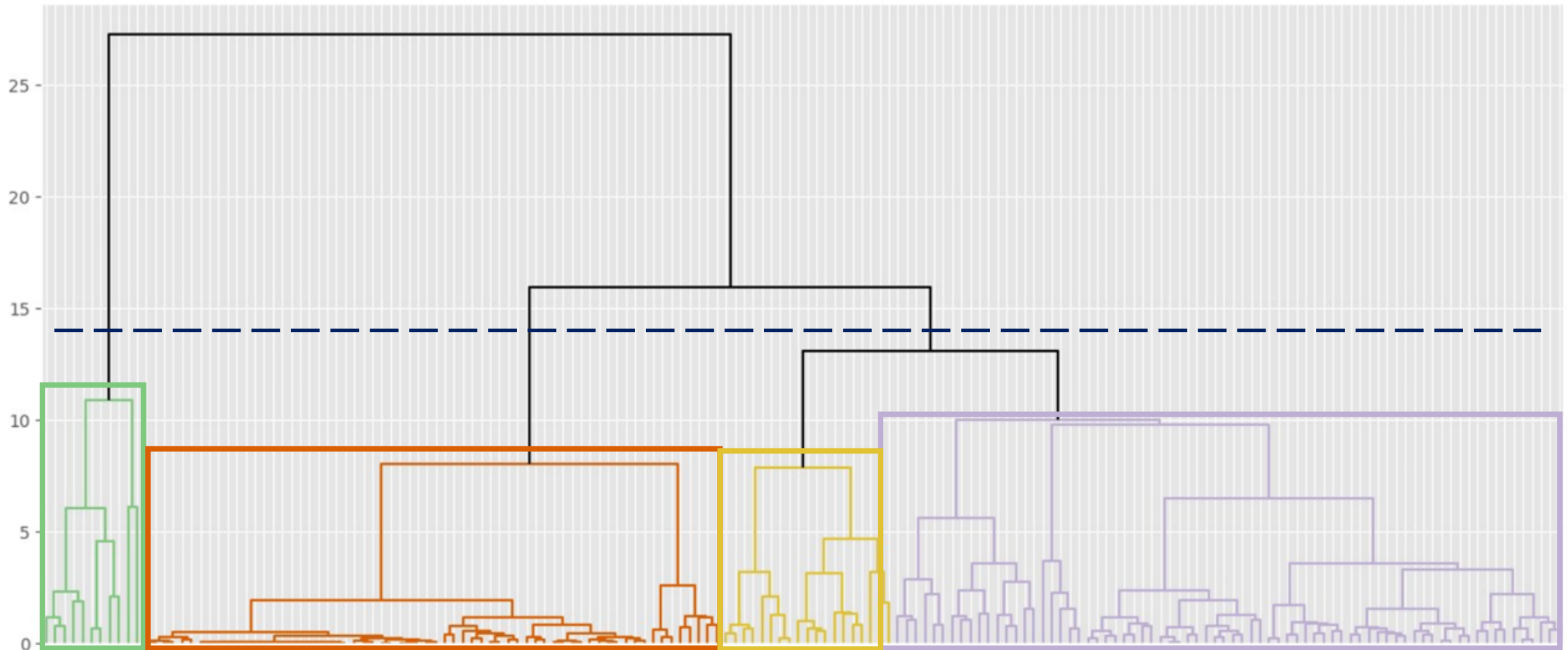


Key Point: 3 to 4 clusters is reasonable

Results – PCA Scatter with Cluster Labels



Results - HCA Dendrogram



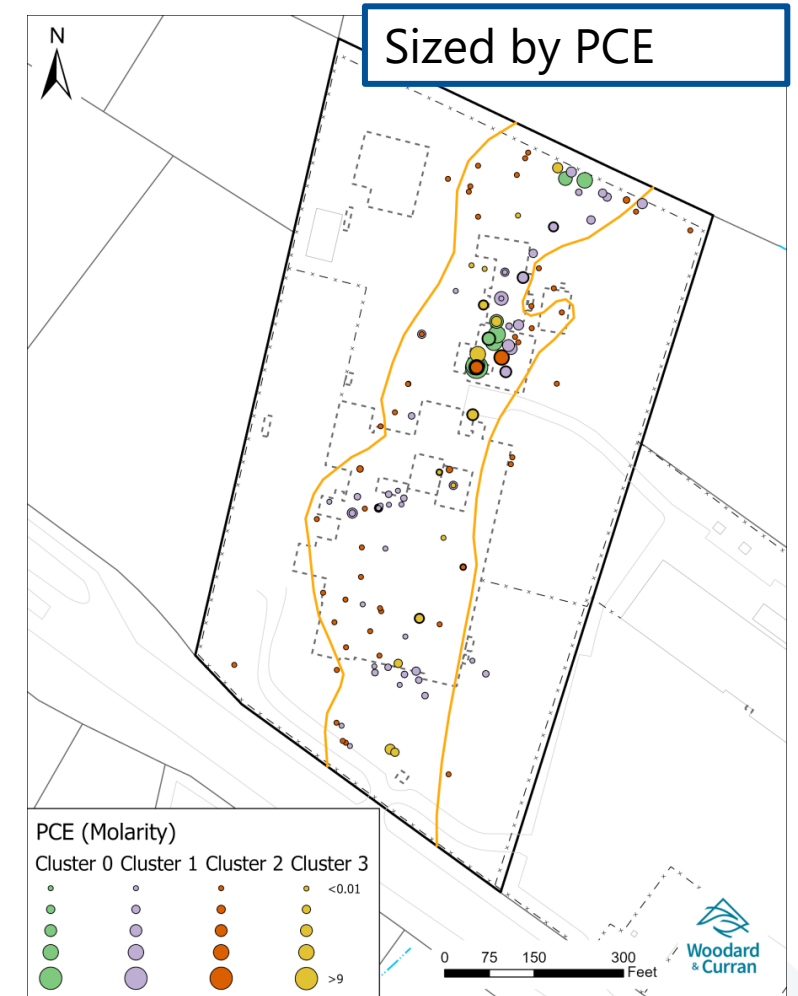
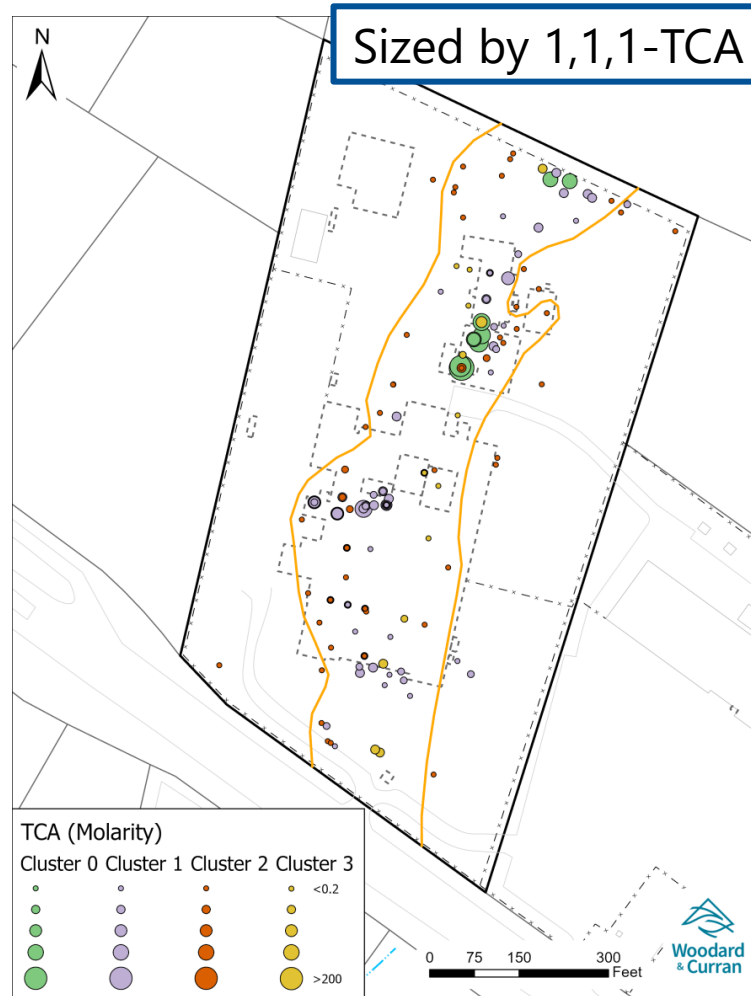
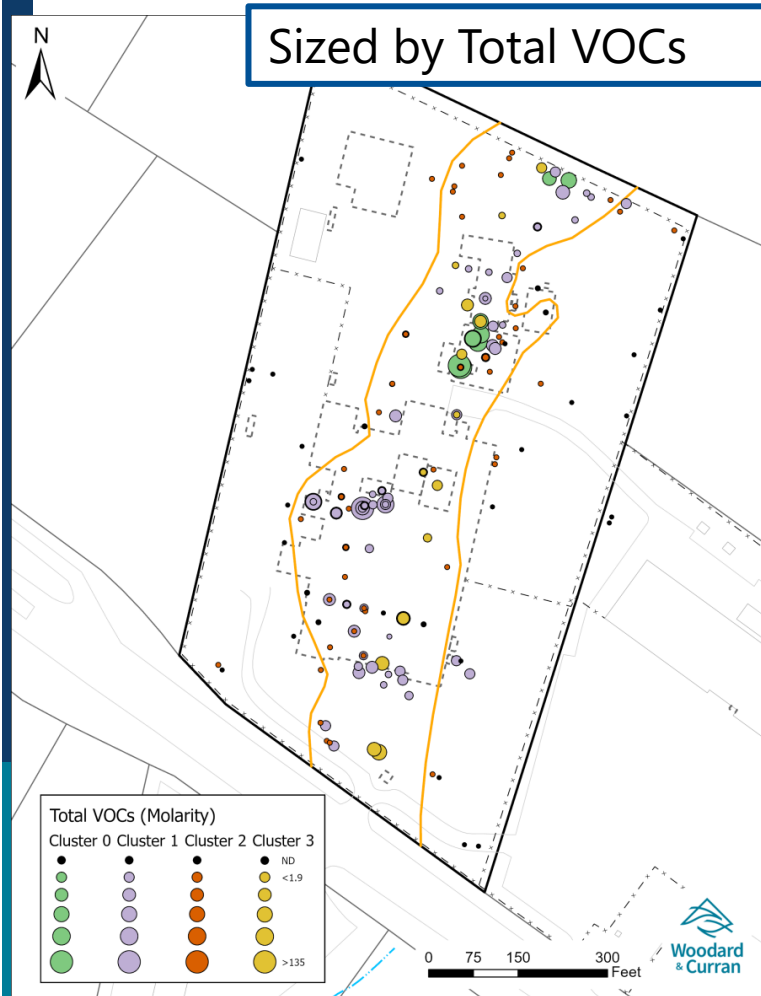
Cluster 0
(1,1,1-TCA)

Cluster 2
(low concentrations)

Cluster 3
(1,4-Dioxane)

Cluster 1
(Mixture of Compounds)

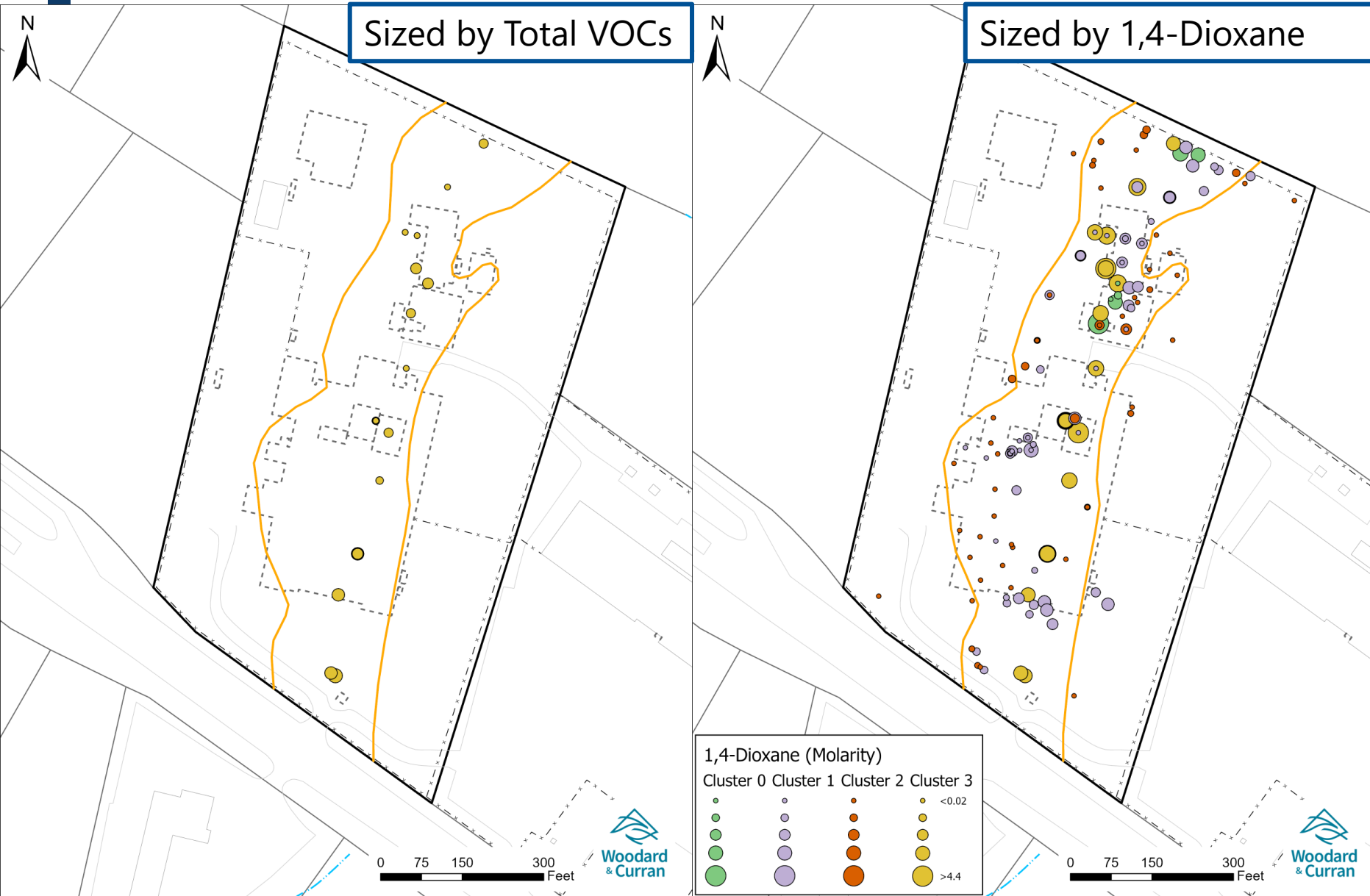
Results – Cluster 0



Key Points:

- 1,1,1-TCA and PCE are highest in Cluster 0.
- Locations limited to northern portion of the Site near the Former Seepage Area
- Wells in Cluster 0 generally exhibit the highest total concentrations

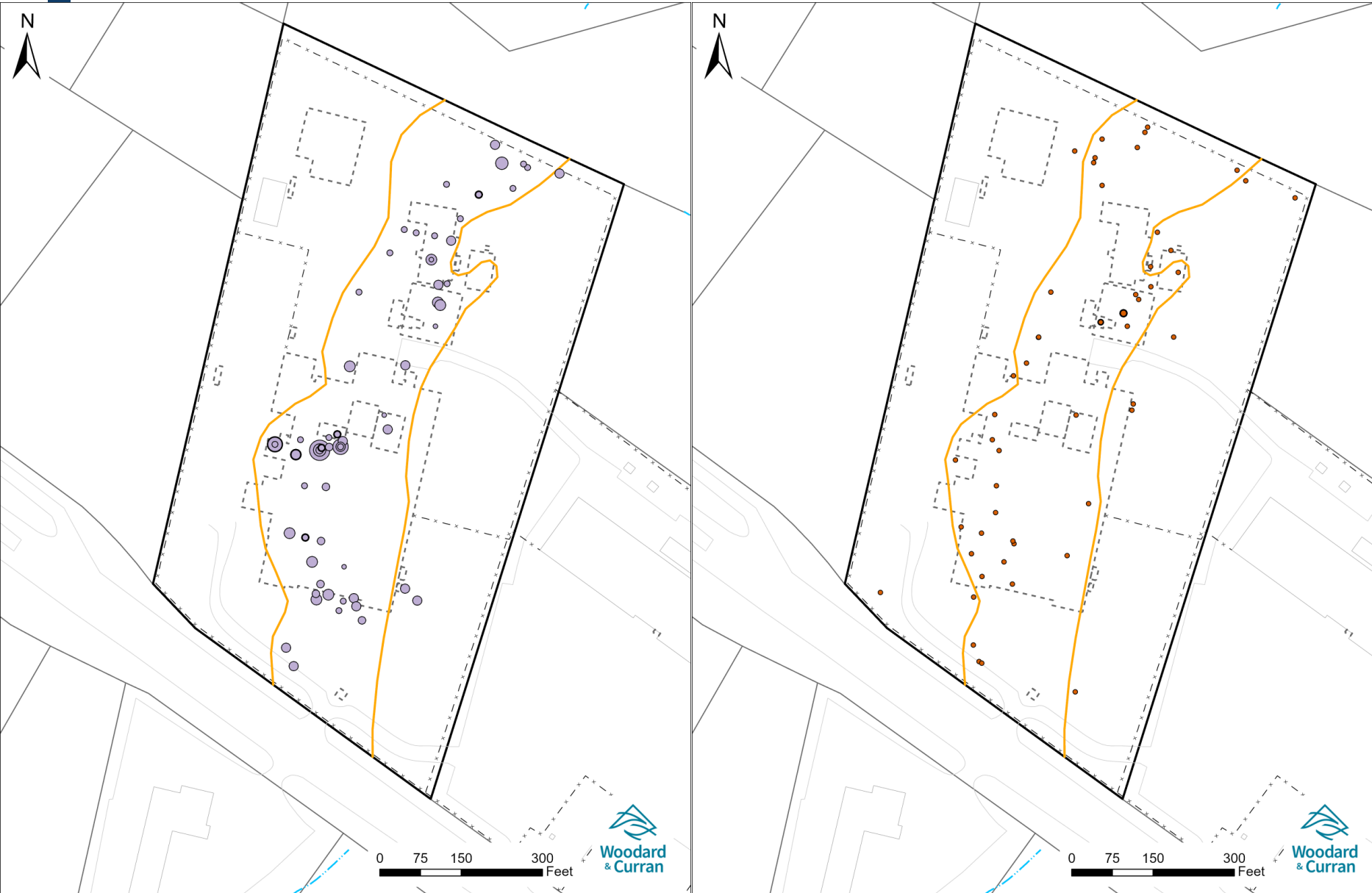
Results – Cluster 3



Key Points:

- 1,4-Dioxane is highest in Cluster 3.
- These wells do not stretch as far east or west as wells in Clusters 1 and 2.

Results – Clusters 1 and 2

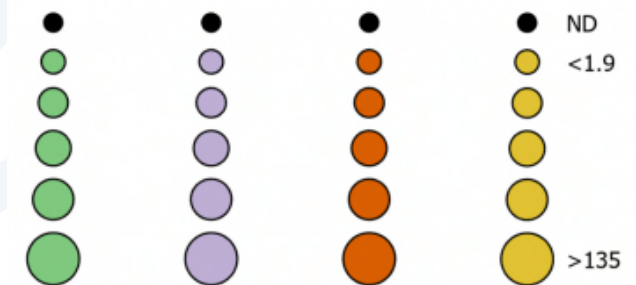


Key Points:

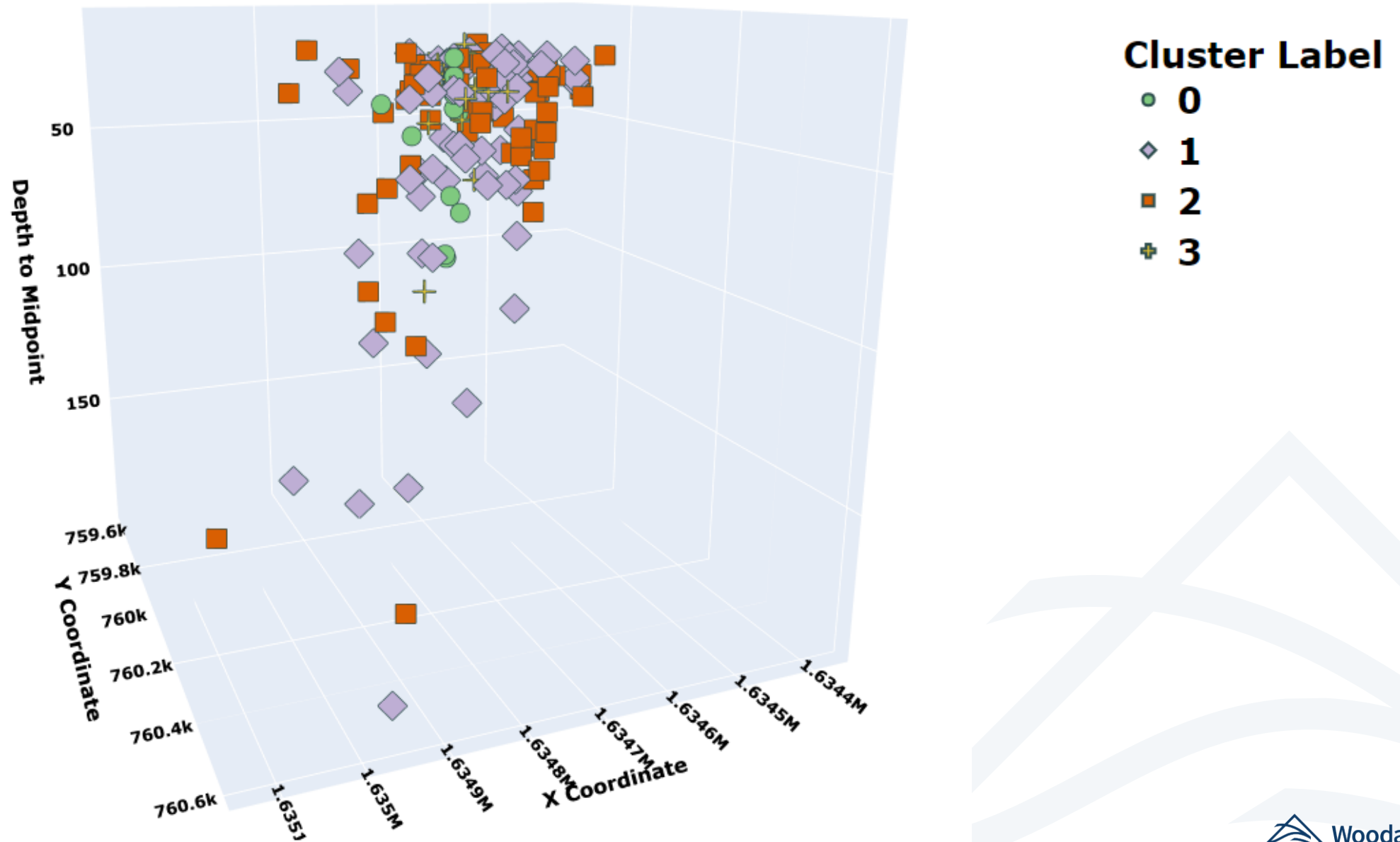
- Clusters 1 and 2 cannot be characterized by individual compounds
- Cluster 2 is characterized by low concentrations

Total VOCs (Molarity)

Cluster 0 Cluster 1 Cluster 2 Cluster 3



X,Y,Z-Coordinates with Cluster Labels



Summary

- PCA and HCA can quantitatively differentiate source and plume areas
- Groundwater samples fall into four clusters:
 - **Cluster 0** = Primarily 1,1,1-TCA, clusters located near the Former Seepage Area. **Does not appear in deep wells.**
 - **Cluster 1** = Mixture of compounds. Includes deeper wells consistent with bedrock behaving as leaky (semi-confined) aquifer
 - **Cluster 2** = Low molarity cluster includes deeper wells
 - **Cluster 3** = Highest molarities of 1,4-dioxane and narrower E-W distribution. **Does not appear in deep wells.**

Relevance

- Successfully reduced a very large data set to focus on a few key variables
- Objective, quantitative differentiation of the plume
 - Two different sources contributing to two different plumes were identified
 - Plume differentiation was vertical as well as lateral
- Used results for targeted site remediation and planning



Questions?

June 5, 2024

vward@woodardcurran.com

