Using Fuzzy C-Means Clustering and Partial Least Squares Regression to Evaluate Sources for Geochemical Changes and Chloride Enrichment in a Calcareous Fen Located in An Urban Environment

ILLINOIS
Illinois State Geological Survey
PRAIRIE RESEARCH INSTITUTE

Richardson, J.J.[1], Schafer, L.A.[1], Miner, J.J.[2], Plankell, E.T.[1], Locke, R.A.[1], Pociask, G.E.[1], and Monson, J.L.B.[1]

[1]Illinois State Geological Survey, [2]Illinois State Geological Survey, Retired

Illinois Nature Preserves Commission

Abstract

Fuzzy C-Means Clustering (FCM) and Partial Least Squares Regression (PLSR) are multivariate methods that have been used in a variety of geochemical and physical settings. FCM has been widely utilized to identify spatial or chemical patterns within a dataset. PLSR is used to create predictive models for variables, based on the response from other variables. These two methods were applied to a long-term water chemistry dataset (n = 763 samples) from a fen to identify potential sources of chemical changes, and to evaluate if those sources can be distinguished by their variances in chloride (Cl) with respect to other analytes.

Bluff Spring Fen (BSF) Nature Preserve is located in the Chicago suburb of Elgin, IL (Figure 1). The site contains calcareous fens and is bordered by Gifford Lake (GL) to the east and an underground mine and business park to the south. GL receives both clean construction debris and roadway runoff, while the mine extracts dolomitic bedrock and infiltrates mine water directly into BSF via an infiltration trench. Water-quality monitoring began at BSF in 2002 to evaluate site dynamics in response to mine development and operation, and adjacent site changes that could affect hydrology and geochemistry.

Five clusters were identified by FCM based on average water chemistry. Clusters 1, 2, and 5 were associated with land uses occurring outside of the preserve boundary (Figure 2; Table 1). Cluster 1 was associated with GL and the eastern portion of the monitoring network, with increased sodium (Na) and Cl, which are attributed to the input of roadway runoff directly into GL. Cluster 2 was associated with the mine water discharge, with increased levels of strontium (Sr), fluoride (F), barium (Ba), and boron (B). Cluster 5 was associated with the central portion of BSF, which exhibited elevated levels of Na, Cl, Sr, and F suggesting mixing of waters both from the mine and GL.

The R^2 values of the PLSR models for Cl from Clusters 1 and 2 were 0.91 and 0.96, respectively (Figure 9). While both models exhibited accuracy in predicting Cl, the analytes and their predictive relationships to Cl were different, suggesting that changes in Cl observed in the fen can be traced to these distinct sources (Figure 15).

Site Location and History

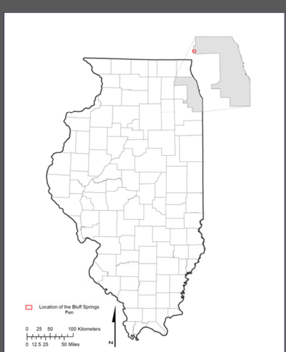


Figure 1. Location of the Bluff Spring Fen Nature Preserve (BSF)

- 2003: Bluff City Materials (BCM) mine opened
- 2004: Water from Gifford Lake was redirected into BSF
- 2006: Infiltration trench was installed to infiltrate mine waters into BSF
- 2010: Road salt was stored on BCM property
- 2013: Discharge from Gifford Lake was rerouted to bypass BSF

Fuzzy C-Means Clustering

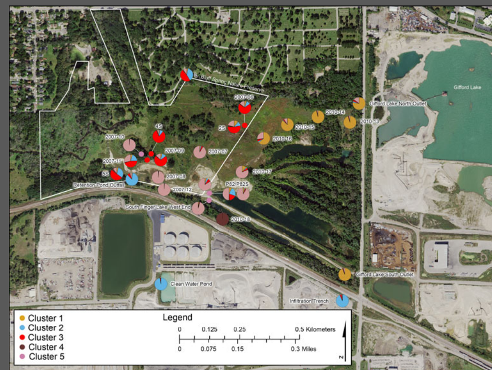


Figure 2. BSF Monitoring Location and Cluster Percentages

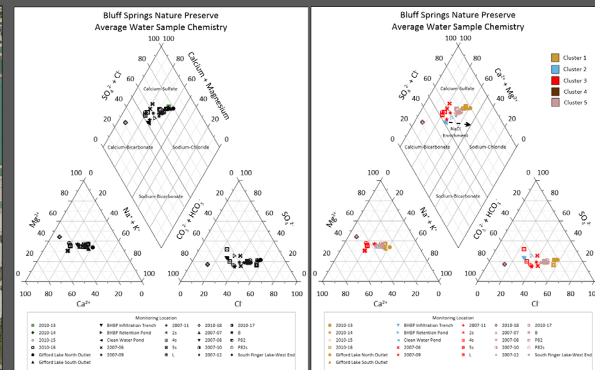
Cluster	Alkalinity	Ba (10^{-3})	B (10^{-2})	Ca	Cl	F (10^{-2})	Mg	K	Si	Na	Sr (10^{-2})	SO_4
1	2.2	0.49	0.7	1.7	5.3	0.81	2.0	0.13	0.19	4.3	0.15	0.95
2	2.9	1.9	2.7	1.8	2.0	2.8	1.7	0.21	0.18	2.4	2.0	0.75
3	3.5	0.61	1.1	2.6	3.0	1.1	2.1	0.087	0.27	2.6	0.27	0.85
4	6.8	0.90	1.5	4.2	1.4	0.59	3.8	0.014	0.12	1.2	0.38	0.86
5	2.6	0.57	1.5	1.7	3.8	1.2	1.9	0.13	0.19	3.4	0.44	0.77

Table 1. Cluster Centers of Average Analyte Values (mmol/L)

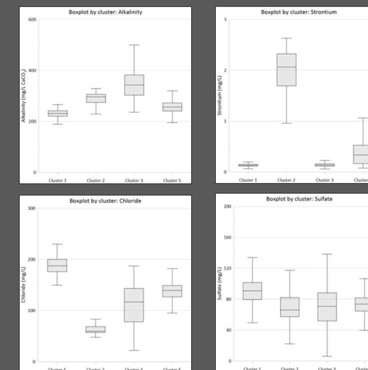
Majority Cluster	Monitoring Location	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)	Cluster 5 (%)
1	2010-13	95	1	1	0	3
	2010-14	96	1	1	0	2
	2010-15	87	1	2	0	9
	2010-16	64	3	5	1	27
	Gifford Lake South Outlet	94	1	1	0	4
	Gifford Lake North Outlet	77	3	4	1	15
2	Infiltration Trench	1	93	4	0	2
	Retention Pond	1	86	7	1	5
	Clean Water Pond	0	97	2	0	1
3	2007-06	5	10	69	2	14
	2007-09	3	12	53	4	21
	2007-11	4	20	45	1	29
	2s	11	12	53	4	21
	4s	2	10	83	1	5
	5s	4	31	51	4	10
	L	3	38	47	2	10
4	2010-18	0	0	0	100	0
5	2007-07	2	2	4	0	92
	2007-08	1	0	0	0	99
	2007-10	1	0	0	0	99
	2007-12	0	0	1	0	99
	2010-17	11	3	5	0	81
	B	4	5	7	0	84
	P82	7	23	19	1	50
	P82s	4	2	2	0	91
	South Finger Lake West End	2	1	1	0	96

Table 2. Monitoring Locations by Majority Cluster with Percentage Inclusions

Fuzzy C-Means Clustering Application



Figures 3, 4. Piper Diagrams of Average Water Chemistry at Monitoring Location by Clusters



Figures 5-8. Boxplots of Select Analytes by Cluster

Partial Least Squares Regression

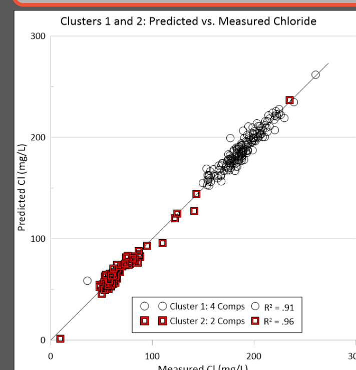
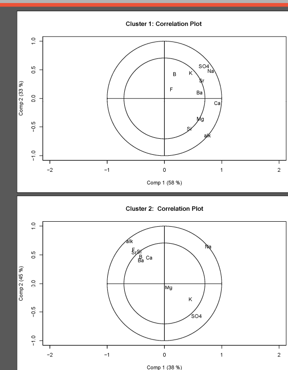
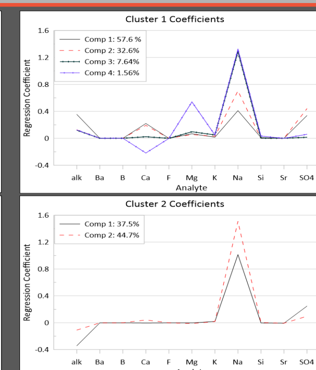


Figure 9. Predicted vs. Measured Cl in Clusters 1 and 2



Figures 10, 11. Correlation Plots, Clusters 1 and 2



Figures 12, 13. Component Coefficients, Clusters 1 and 2

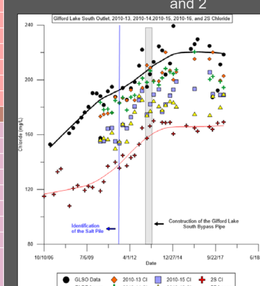


Figure 14. Measured Chloride in Cluster 1 and Well 2s

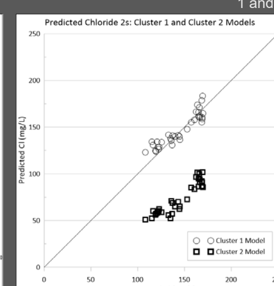


Figure 15. Modeled vs. Measured Chloride in Well 2s

Conclusions

This study has shown that Fuzzy C-Means Clustering and Partial Least Squares Regression can be used together to identify potential urban sources of chemical influence and to track observed changes in a complex geologic system.

Acknowledgments: This work was partially funded by the Illinois Nature Preserve Commission and the Illinois Division of Natural Resources – Office of Resource Conservation. Data collection was performed by numerous staff from the Illinois State Geological Survey and GZA Geoenvironmental Inc. on behalf of Vulcan Materials Company. Water samples were analyzed at the Illinois State Water Survey Public Service Laboratory.